The Role of Output Tasks in Second Language Acquisition

博士論文

The Role of Output Tasks in Second Language Acquisition
(アウトプットタスクの第二言語習得における役割)

Saruul Borjigin
(包　薩如拉)

2014 年

Table of Contents

Chapter 1 Introduction

1.1 Overview of the chapter

This chapter provides an overview of the theoretical framework for this dissertation research in the first part. Then it goes on to present the significance of the study along with its purpose. In the final part, the organization of the whole dissertation is outlined.

1.2 Attention and noticing in second language acquisition

There is now a general consensus within the second language acquisition (SLA) field that attention to certain linguistic features (form) in input is necessary for learning[1] to take place (see Schmidt, 2010 for a review). It is claimed in the Noticing Hypothesis (Schmidt, 2001) that "people learn about the things that they attend to and do not learn much about the things they do not attend to" (p. 30). Therefore, how to draw learners' attention to target linguistic features effectively so as to promote the learning of those forms is the current interest of researchers as well as language teachers. This led to the proposal of focus on form (Long, 1991) or form-focused instruction (Spada, 1997). While Long defined his proposal (focus on form) as something that "overtly draws student's attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication" (pp.45-46), form-focused instruction (FFI) is defined as "any pedagogical effort which is used to draw the learners' attention to language form, either explicitly or implicitly" (Spada, 1997, p.73).

Spada (1997) claimed that her definition of FFI is essentially different from the focus on form. According to her the difference between these two terms is that "… focus on form is restricted to meaning-based pedagogical events in which attention is drawn to language as a perceived need arises rather than in predetermined ways. … FFI...refer to pedagogical events which occur within meaning-based approaches to L2 instruction but in which a focus on language is provided in

[1] The term learning and acquisition are used interchangeably throughout this thesis. There is a debate about the difference between these two terms in SLA field, with some researchers (e.g., Krashen, 1985) distinguishing them as two different systems: learning takes place consciously and acquisition unconsciously. However, in contemporary SLA terminology no such distinction is typically upheld (Ortega, 2009, p.5).

either spontaneous or predetermined ways." (p. 73). Therefore, the present study follows the FFI definition in operationalization of the two output tasks. According to Skehan (2003) different approaches that aim to draw the learner's attention to linguistic forms in input, "agree the role noticing play in the language learning, whether this is through input or output" (p. 2). While noticing through input (e.g., textual enhancement; input flood; input processing) has been substantially researched (see Lee & Huang, 2008 for a review), noticing through output (i.e., noticing function of output hypothesis by Swain, 1985; 1995; 2005; see also Muranoi, 2007) has been relatively less researched.

1.3 Integrated model of second language acquisition

So how exactly does noticing affect learning? In explaining this inquiry it is necessary to have a model that explains how language learning takes place. Among several general learning models that have been proposed by different researchers (e.g., Ellis, 1993; Gass, 1988) on the process of the acquisition of linguistic knowledge in a second language[2], the model proposed by Gass (1988) has been widely implemented in SLA studies because it is considered to provide a detailed explanation of the each stage learner goes through to finally achieve language acquisition. The model includes five stages that learners go through to acquire the linguistic knowledge of L2. It begins from ambient speech to apperceived input to the final stage of output. It is presented in following figure.

[2] The term second language (or L2) in this thesis refers to any language that has been learned after the learners' first language. Therefore, although for the participants in the present study, English is actually their third language, I did not distinguish between L2 and L3, because this alone can be a research topic and thus beyond the scope of the present study.

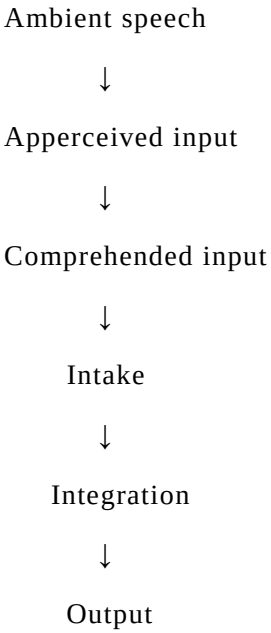


Figure 1.1 Gass's model of second language acquisition (from Gass 1988: 200)

Gass (1988) explains this model in following way. Ambient speech refers to the second language input the learners are surrounded by. However, not all of this input data is utilized by the learners. Only certain parts of this data are going to pass through to the next stage, which is defined as apperceived input. The factors which are influencing this filtering process can be the frequency of the linguistic information, affective variables (e.g., motivation, attitudes), prior knowledge and attention. Anyway, it is clear that in order for initial ambient speech to pass into apperceived input stage, some aspects must be noticed or attended to.

Attention, in Gass's term "is what allows a learner to notice a mismatch between what he or she produces/knows and what is produced by speakers of the second language" (p. 203). In other words, the information which is entered into apperceived input stage is also what enables learners to notice the gap between their interlanguage (IL) and target language (TL).

Once a certain amount of data is passed into apperceived input, they go onto make their way into comprehended input stage. By comprehension, Gass refers to the analyses learners have done to the input at this stage. She believes the analyses are multi-staged, which range from level of semantics to linguistic structural analyses. She points out that the level of analysis done to comprehended input is important for the subsequent intake stage to take place. Intake is defined in

this model as a process of mental activity which mediates between input and grammar. And it is these mental activities that enable new linguistic knowledge to integrate into learners' already existing IL system. This integration process can: (1) confirm the current hypothesis learners hold in their IL system and thus facilitate the integration of new piece of knowledge into their developing IL system; (2) or if the knowledge is already part of what learners know, then this process can serve as reconfirmation and rule strengthening in IL development; (3) or reject some of the existing hypotheses and therefore make learners seek further relevant input for further intake. The final stage in the model is output. In defining the role output play in language acquisition, Gass joins Swain (1985) on claiming that 'comprehensible output' is necessary for successful acquisition to take place. She then explains how output functions in the model as : "using the language forces learners to make sophisticated analyses of grammar, a factor which is important to in moving the learner from comprehended input to intake "(p. 210). It can be inferred from this statement that in fact this model does not end at output stage. This final stage also provides feedback far back into the input stage and keeps the cycle going on.

In sum, what this model is suggesting is that learners do not learn all input they are exposed to. Only a certain amount of the ambient speech can pass into intake. And once they reach intake, they will be either integrated into the learner IL system, or it will make learners seek further input to derive the knowledge they need, which again leads them to the input stage. Based on this knowledge they will finally be able to use the language. Then, upon using the language, they may further notice some mismatches between their IL and TL and this again leads to the input stage of the model.

In summarizing how the model functions, she concludes by acknowledging the important role attention plays in the whole process in the following sentence:

> "In sum I am suggesting a major role for what I have called apperceived input, determined to a large extent by selective attention. Without selective attention grammar development does not take place. In other words, a first step in grammar change is the learner's noticing (at some level) a mismatch between the ambient speech and his or her own organization of the target language" (Gass, 1988, p. 212).

Put simply, to develop their IL system, learners need to notice the gap between their IL and TL. One such way to make them aware of this gap is 'comprehensible output', which was mentioned by Gass in explaining the output stage in her model. This notion was first introduced by Swain (1985) whose observation of the learners from a French immersion program led her to question the role of 'comprehensible input' in learning. Until then, it was believed that 'comprehensible input' (Krashen, 1982; 1985) was the only one necessary and sufficient condition for language acquisition to take place. Swain and her colleagues found that English speaking children who were enrolled in the French immersion program displayed near native competence in listening and reading comprehension abilities, but not in their speaking and writing abilities. This led to her proposal of the role output plays in language learning, initially known as 'comprehensible output' (Swain, 1985) and later the 'output hypothesis' (Swain, 1995; 2005).

1.4 Output hypothesis

In Swain (1995, 2005), she proposed four hypothesized functions of output in language learning. They are:

(1) the noticing/ triggering function;

(2) the hypothesis testing function; and

(3) the metalinguistic function.

Swain (1995, 2005) explained these functions in following ways.

First, through producing the target language, learners may find out the gap between what they want to say (or write) and what they actually know. This may lead them to realize their own linguistic problems, that is, what they don't know or only know partially. This noticing of the gap in their linguistic knowledge may trigger their paying attention to the linguistic knowledge they need from the relevant input. In so doing, learners may either generate new knowledge or consolidate their existing knowledge. For example:

Non-native speaker: And in hand in hand have a bigger glass to see.

Native speaker: It's err. You mean something in his hand?

Non-native speaker: like spectacle. For older person.

Native speaker: Oh aha I see you mean a magnifying glass, right?

Non-native speaker: yeah, that's it.

(Mackey, 2002, pp, 389-390, modified by the author)

In this example, the non-native speaker wanted to say magnifying glass, but did not know how to say it (i.e., noticing the gap). So when the native speaker provided him the right word, he picked that up (i.e., seeking answer form the relevant input).

Second, through producing the target language, learners may have the chance to test the hypothesis they hold in IL. For example:

Non-native speaker: And in hand in hand have a bigger glass to see.

Native speaker: It's err. You mean something in his hand?

Non-native speaker: like spectacle. For older person.

Native speaker: Mmm, sorry I don't follow, it's what?

Non-native speaker: In hand have he have has a glass for looking through for make the print bigger to see, to see the print, for magnify.

Native speaker: He has some glasses?

Non-native speaker: Magnify glasses he has magnifying glasses.

Native speaker: Oh aha I see a magnifying glass, right that's a good one, ok.

(Mackey, 2002, pp, 389-390)

This example shows the process of how the learner tests the hypothesis exists in his IL. At first he struggled to produce the exact word, but when finally he came up with the word 'magnifying glass', the native speaker provided the confirmation feedback to his hypothesis such as 'Oh aha I see 'or 'right that's a good one'. Therefore, through such opportunities to produce output (say or write) and receive feedback from a native speaker or language teachers may enable learners to either consolidate their existing knowledge or seek a solution in the input (spoken or written form).

Third function of output is the metalinguistic function or reflective function. This function in

Swain's term is: "Under certain task condition learners will not only reveal their hypothesis, but reflect on them, using language to do so" (Swain, 1995, p. 132). Put simply, when learners try to reflect on a particular linguistic form in the L2, using some metalinguistic knowledge they already have about that form, they may benefit from such experience. For example:

> Student : I need many practice in English to improve my fluency.
> Teacher : Why do you use many?
> Student : Oh, right. I should use *lots of* instead of *many*, because *many* can only be used before countable nouns.
>
> (Created by author)

Swain (1995) claimed that this kind of "using language to reflect on language" output, (i.e., metalinguistic reflection) may allow learners to control and internalize the linguistic knowledge in L2. In her later work drawing on social cultural theory, Swain relabeled the output into speaking, writing, collaborative dialogue/verbalizing.

The second and the third function (i.e., hypothesis testing and the metalinguistic function) have widely been researched in the fields such as feedback and subsequent modified output (see Swain 2005 for a review) languaging and collaborative dialogue (Swain 2005; see also Swain & Watanabe, 2013). However, despite the importance of the role of noticing in SLA, only limited numbers of studies have been conducted to investigate the noticing function of the output hypothesis up to date (e.g., Izumi, 2002; Uggen, 2012). Therefore, the present study sets out to investigate the noticing function of the output hypothesis and attempts to address such research questions as whether or not the output actually triggers noticing and whether such noticing facilitates the learning of the target form. In Chapter 2, the previous studies on noticing/triggering function of output are reviewed.

1.5 Significance and the purpose of the study

The purpose of this study is to examine: (1) whether providing learners with the opportunity to produce output actually better facilitates the noticing and acquisition of the target forms in the

input passage than asking learners to read the passage for comprehension only; (2) based on the results and findings of the study to explore the relationship between output triggered noticing and acquisition; (3) as well as the possible roles task type and individual differences might play in predicting the effects of output. Findings from this study are expected to provide theoretical implications for SLA literature and pedagogical implications for L2 classroom, especially the ones in EFL situation where typically the class sizes are big and still exclusively focusing on grammar translation method. The two FFI output tasks utilized in the present study can easily (by easily, I mean without much modification) be introduced and applied to such classrooms and are expected to contribute to both teaching and learning of foreign languages.

1.6 Organization of the dissertation

The chapters of this dissertation are organized in following ways. Chapter 1 introduces the theories and hypotheses that led to and framed the present study along with the purpose and significance of the study. Chapter 2 reviews the empirical studies and addressed the remaining problems, then sets the specific research questions of this study. Chapter 3 describes the methods that I used in the study in order to answer my research questions. This part includes the testing materials, tasks, treatment procedures and the rationales for the chosen materials as well as the criteria for scoring and analyzing the data. Chapter 4 presents the results of analyses done in Chapter 3. In Chapter 5, the results and findings from Chapter 4 are interpreted with reference to theories and empirical studies reviewed in Chapter 1 and 2. Finally, Chapter 6 summarizes the major findings of the study, then goes on to acknowledge some limitations of the study and offer some directions for future studies. It concludes with some pedagogical implications. Testing materials and treatments packages will appear in the appendices.

Chapter 2 Literature Review

2.1 overview of the chapter

In this chapter previous studies on output hypothesis are reviewed. Then after summarizing the findings of previous studies some of the remaining problems are pointed out. Finally, the specific research questions of this study are presented.

2.2 Previous studies on noticing function of output hypothesis

Izumi carried out a series of studies (Izumi 2002; Izumi & Bigelow, 2000; Izumi, Fujiwara, Bigelow & Fearnow, 1999) to address: (1) whether output promotes noticing of linguistic form; and (2) whether output results in improved performance of the target form.

The first study was conducted by Izumi and his colleagues (Izumi et al., 1999) on effects of output triggered noticing and acquisition of past hypothetical conditional structure in English. They included two output treatments in their study. During Phase1 treatment, they asked the participants to first read the input passage and reconstruct, then go on to repeat the same procedure again. The Control group in this phase just read for the sole purpose of comprehension. After the Phase1 treatment, they administered posttest 1, and then, during the next week, they carried out Phase 2 treatment in which they first asked participants to write an essay on a given topic then provided a model written by a native speaker and after that asked the participants to carry out the essay writing on the same topic again. After that they assigned posttest 2. The treatment procedure is illustrated in the following figure.

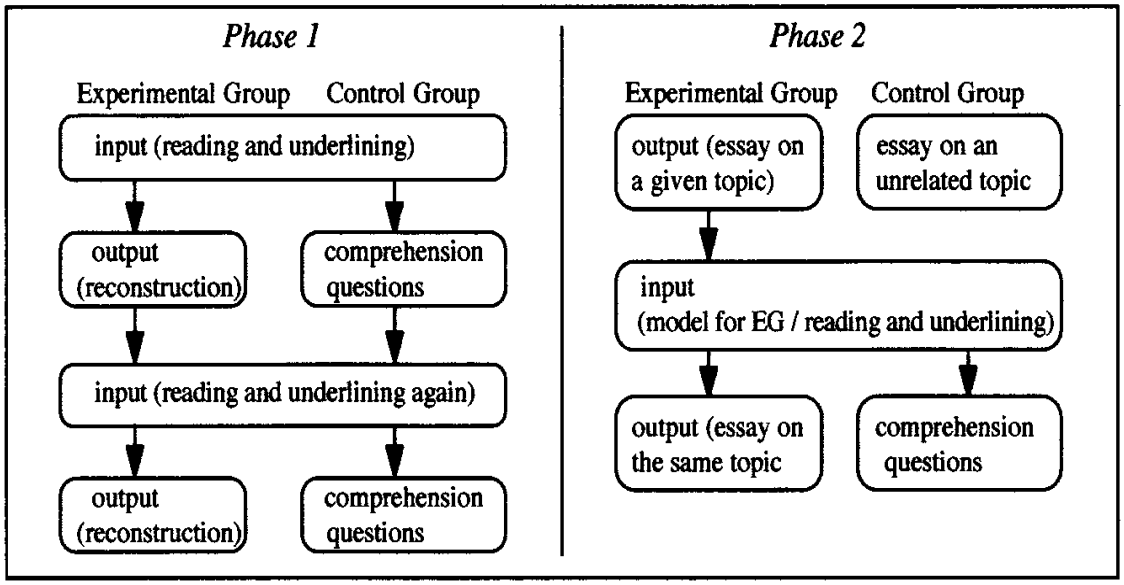


Figure 1.2 treatment procedures in Izumi et al., 1999 (p. 428)

In both Phase 1 and Phase 2, the noticing was captured through underlining activities carried out by the participants in each input session. The acquisition was measured by their performance on a grammaticality judgment test and picture-cued production test.

In terms of noticing, the findings showed that the reconstruction task was more effective in drawing learners' attention to target form and also demonstrated greater immediate incorporation of the target form than essay writing. On the other hand, with regard to acquisition, essay writing turned out to be more effective, because of the improved performance of learners in posttest 2. These findings imply that output (i.e., reconstruction and essay writing) promote noticing of the target form and thus better facilitates second language acquisition. However, it should be noted that, in terms of noticing, output group did not outperform the control group. Although in posttest 2, output group improved from pretest to posttest picture-cued production test scores, and outperformed the control group, they did not achieve the same on their performance on the grammaticality judgment test.

Drawing on the above findings, they concluded that although this study failed to support the noticing function of output hypothesis, since the output group did not show better noticing than the control group, the study did come up with a partial support for the facilitative effect of output on production knowledge of the target form.

However, because the same participants from Phase 1 treatment carried out Phase 2 as well, it

was difficult to argue task type effect and besides whether the improved performance in posttest 2 was facilitated by essay writing task alone or by combination effect of both tasks was uncertain.

Then in Izumi and Bigelow (2000), they conducted a follow up study to Izumi et al. 1999. This time they switched the sequence of Phase 1 and 2 (i.e., essay writing first and reconstruction next) to address noticing and acquisition of past hypothetical conditional structure along with the differential effect of task sequencing. Again, the noticing was measured by the underlining scores and acquisition was measured by multiple-choice recognition test and picture-cued production test. As it turned out, in terms of noticing, the output group failed to outperform the comparison group. In terms of acquisition, the results were of partial support to research question two, because the output group improved from their production posttest 1 to posttest 2. However, the comparison group manifested the same pattern as did the output group and there was no difference between the groups.

On the other hand, based on the findings from the performance of the output group on their immediate uptake in subsequent input, Izumi & Bigelow (2000) pointed out that providing learners with extended opportunities to produce output and relevant input was crucial in improving the learners' use of target language. In addition, by further analyzing the interview data and production data during the treatment stage, they found that not all participants were able to realize that their IL grammar was problematic. This led to their suggestion for future studies to explore different task type which could make learners be aware of the gap between their IL and TL and different linguistic features, because it is possible that some linguistic features are not easy to be noticed.

Uggen (2012) replicated this study (i.e., Izumi & Bigelow, 2000) and investigated the effects of output on subsequent input noticing and production of target form. Only, this time she included complexity of linguistic structure (i.e., present hypothetical conditional vs. past hypothetical conditional) as one of the factor that might mediate the effect of output on noticing. Namely, she hypothesized that more complex and difficult structure (i.e., past hypothetical conditional) may trigger greater amount of noticing. The noticing was measured through both underlining and stimulated recall. The acquisition was addressed through picture-cued production tests.

She found that the output group who were asked to produce past hypothetical conditional (i.e., more complex structure) outperformed comparison group in both noticing scores and production

posttest scores. Interestingly, the output group who produced present hypothetical conditional (i.e., less complex structure) failed to do so. Therefore, this study has only provided a partial support for output hypothesis.

Drawing on these findings she argued that structural complexity may not necessarily cause cognitive overload, on the contrary, it plays a positive role in promoting noticing. In contrast, the simple structure may appear not so salient for learners to notice. This, she argued, indicated the relationship between cognitive demand and noticing and suggested the necessity of further investigation.

A recent study by Borjigin, Suzuki and Itagaki, (2013) examined the effect of output on noticing and acquisition of the English passive voice. In addition, they also investigated the differential effect of cognitive demands of tasks. The acquisition was measured through multiple-choice test. They carried out two experiments. In first experiment, they asked learners from the output group to read and reconstruct the input passage and they repeated the same procedure twice. The comparison group just read the passage and completed a comprehension check question. As a result, the comparison group improved but not the output group on the posttest. In the following experiment, they only included the output group with the same procedure and same testing material. However, this time they modified both the input passage and the output task. Namely, they simplified the input and then modified the reconstruction task into a cloze reconstruction task. They asked learners to only reconstruct the target form (i.e., passive forms) and any other irrelevant linguistic forms were provided. As a result, the group improved on their posttest performance.

This result led them to argue that a cognitively less difficult task may better facilitate the learning of the target form. However, since they also modified the input passage, whether this improved performance was resulted from the output task or the input or even, the combination of both was uncertain.

These studies were all comparing the effect of output on noticing and subsequent learning with that of comprehension as the only condition. Izumi (2002) set out to explore if output (i.e., internal attention drawing device) is more effective than input enhancement (i.e., external attention drawing device) on noticing and acquisition of the English relative clause.

The noticing was measured through note taking and acquisition was addressed by two tests on

receptive knowledge (i.e., an interpretation test and a grammaticality judgment test) and another two on productive knowledge (i.e., a sentence combination test and a picture-cued sentence completion test). The treatment procedure is illustrated in figure below. + and - indicate with or without that opportunity. Thus, output group fell into two sub-groups: with or without input enhancement (i.e., +O + IE; + O - IE) and similarly input only group also could be divided into two sub groups: with or without input enhancement (i.e., -O + IE; -O - IE).

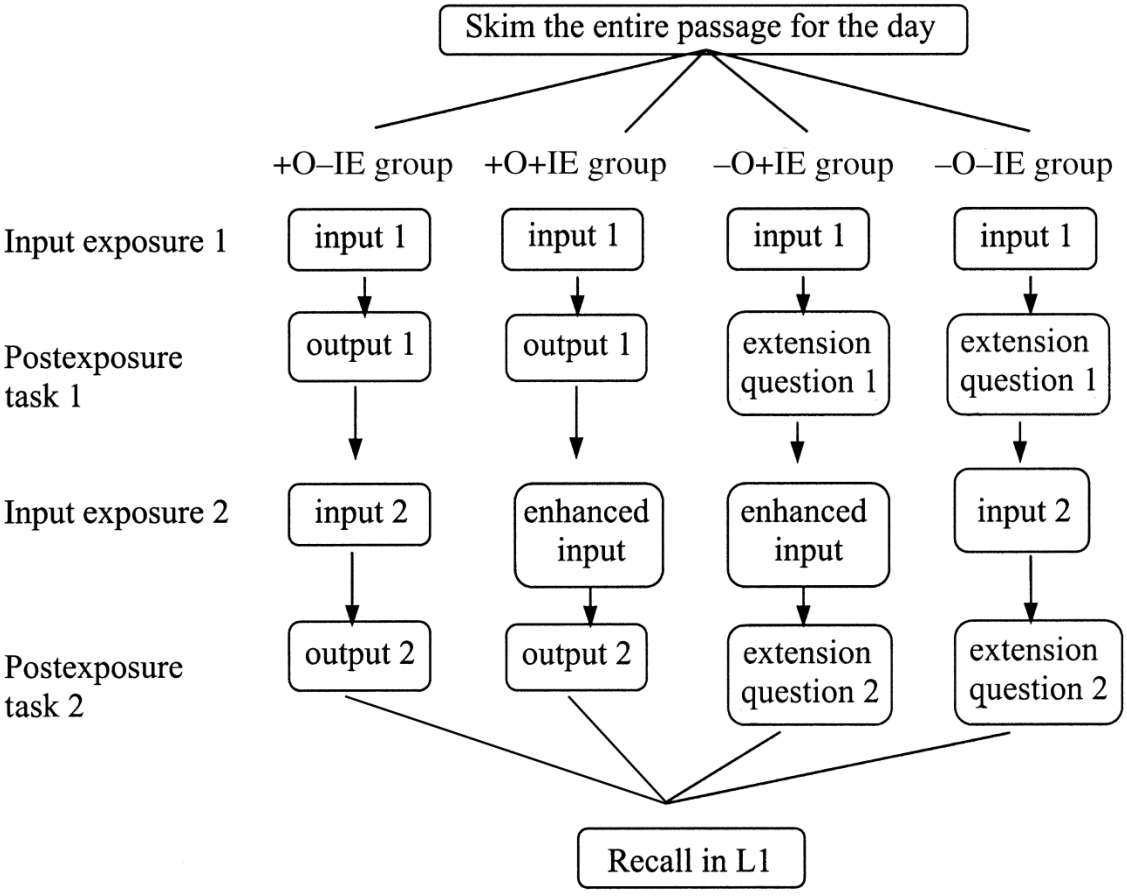


Figure 1.3 treatment sequences in Izumi, 2002 (p. 553)

The findings showed that while input enhancement treatment triggered more noticing of the target forms, it was the output treatment that promoted acquisition of the target form. He then went on to argue that input enhancement succeeded in drawing learners attention to form, but did not necessarily encourage further cognitive processing that may be important for learning to take place. According to his speculation, input enhancement may have succeeded in drawing learners'

attention to form, but did not lead to their noticing of the mismatch between their IL and TL. Output, on the other hand, promoted both the noticing of form and the mismatch, and thus promoted ultimate learning results. Drawing on all these observation, Izumi (2002) suggested that depth and level of noticing was more important than amount of noticing.

2.3 Task difference and output triggered noticing and acquisition

As has been repeatedly pointed out by the empirical research both in output hypothesis and FFI literature, the differential effect of task type has not been thoroughly studied and worth further research. The findings from such research should contribute to the construction of theory and pedagogical application. Especially in output hypothesis literature, there were only limited number of studies that examined the role task types play in mediating the effect of output on noticing and acquisition. One such study was carried out by Song and Suh (2008) to investigate the differential effect of two output tasks (i.e., reconstruction vs. picture-cued writing task). They found that picture-cued writing task was more effective in promoting noticing compared to reconstruction task. However, no differential effect of task type was confirmed in terms of acquisition. Ghari and Moinzadeh (2011) also attempted to compare the different effect of reconstruction task and picture-cued writing task and found similar results. Song & Suh (2008) claimed that reconstruction of the whole text may attract learner attention to some other irrelevant forms other than the target form alone. This suggestion accords with what Borjigin et al. (2013) found that cognitively less demanding output task (i.e., cloze reconstruction) is more facilitative than high demanding task (i.e., text reconstruction).

About operationalizing tasks, one option Robinson et al, (2012) proposed drawing on findings from Ackerman & Ciancolo, (2002) is to match the task character and the ability the learner brings to it. In their terms "…by delivering tasks having the same or similar characteristics, and research the ways these task characteristics make demands on the abilities learners bring to them, and the consequences of these task characteristic -ability determinant coordinates for success or failure in learning " (p. 261).

This direction motivates the present study to explore reconstruction cloze task and reconstruction editing tasks. They are both reconstruction tasks, based on same input passage, and

with same linguistic features to be reconstructed, only through different ways, one is by filling in the gap form, another by error correction, which is likely to draw more on metalinguistic knowledge (Sheen, 2007).

2.4 Individual differences, noticing and acquisition

It has been mentioned in the previous chapter that SLA theories and empirical research have come to a general understanding that L2 learners need to focus their attention on the forms of the target language in some ways. Then a concern may arise along with it as to whether some learners are better at noticing forms than others and if so what makes them different.

While researchers tended to study individual differences such as language aptitude (Robinson, 2005a; Skehan 1998) and motivation (see Dörnyei &Ushioda 2010 for a review) and their roles in SLA achievement in general, in recent years however, there is a new tendency in individual difference research to focus on the link between certain cognitive ability (e.g., working memory, language aptitude) and the language learning process (see Robinson et al, 2012 for a review).

For example, according to Skehan (1998), in adult language learning, different aspects of language aptitude may operate differently at different stages of the language learning. Namely, language analytical ability is involved throughout the whole process of language learning; phonemic coding ability plays a major role only in early stages; and memory ability is involved in all stages but it is enhanced especially in later stages. This assumption was proved to be true in a recent study conducted by Mukoyama (2009; see also Mukoyama, 2013) which investigated the acquisition of L2 Japanese by Chinese learners.

In his later work (Skehan, 2002), he also suggested that a different component of aptitude may relate to the four macro stages in learning: noticing (for example, phonemic coding ability), patterning (for example, language analytical ability), controlling (for example, memory retrieval process), and lexicalizing (for example, memory abilities).

A number of studies have attempted to address this issue. For example, Mackey & Sachs, (2011) and Mackey, Philp, Egi, Fujii & Tatsumi (2002) attempted to address whether greater working memory capacity better facilitates the ability to notice and use the negative feedback provided in interaction. The results supported their assumption that learners with greater working

memory capacity did benefit from this type of feedback and also demonstrated better noticing and learning. Similarly, Sheen (2007) examined the extent to which language analytical ability mediated the effects of two different focused written corrective feedback (direct only correction vs. direct metalinguistic correction) on the acquisition of articles in L2 English. Her findings showed that both types of feedback proved to be beneficial for acquisition of the articles, and the direct metalinguistic correction feedback showed strong correlation with language analytical ability. In other words, acquisition of articles is more effective when direct metalinguistic correction is provided and the learners have higher language analytical ability.

Ranta (1998) on the other hand, focused on the grammatical sensitivity and its role in language learning. She perceived language analytical ability (grammatical sensitivity and inductive language learning ability) as the learner's ability to focus on form and hypothesized that learners with higher level of grammatical sensitivity would have advantage in some aspects of L2 learning. The grammatical sensitivity test was administered in the learners' L1, French. She found that the scores of grammatical sensitivity scores associated with higher stages of grammatical development. These findings support that there is a link between individual's cognitive ability and concept of noticing and subsequently acquisition.

So far, only few studies have been conducted to see the relationship between individual differences and the possible role they play in output triggered noticing and subsequent learning. One such study was Hanaoka (2007). He examined if learners' proficiency level had mediating effect on the effects of output (picture-prompted composition writing) on the noticing of the linguistic features in subsequent native speaker model. Learners were of two different language proficiency levels (advanced vs. intermediate). The results showed that by producing output (i.e., composition writing), the learners discovered their own linguistic problems and looked for solutions from the subsequent model and incorporated them into their subsequent revisions. However, while learners from advanced level did show a tendency to notice more problems than the less proficient group, the incorporation scores of the noticed forms into their subsequent revision were not significantly different from that of less proficient group. Besides, there was no control group in this study. What this study demonstrated was that output activity (i.e., composition writing in this study) provide learners with opportunity to notice the linguistic problems in their IL system. However, whether this noticing effect was influenced by their

language proficiency level was not confirmed. This might have been due to the fact that he did not divide the two groups based on any test scores. He just picked up one group of students who were considered advanced in their department and the other group of learners who were considered low intermediate from another department in the same school.

Another study that took into account the role of language proficiency in light of output hypothesis was conducted by Suzuki, Itagaki, Takagi, & Watanabe (2009). They investigated the effect of output on subsequent input through asking their subjects to carry out a picture description task. The participants in their study comprised of two groups: low-intermediate proficiency group (first year high school students) and high-intermediate proficiency group (third year university students). Each group was then further divided into two subgroups: output group and non-output group.

Results showed that the output group outperformed the non-output group in their recall task. This indicates that noticing triggered by output facilitated the learners' attention to relevant linguistic information in the subsequent input and thus promoted learning. With regard to the role of proficiency level in such effect, they found that the high level group outperformed the low level one regardless of the treatment type (i.e., output or non-output). They claimed this was because of the limited memory capacity of the participants in low level group.

Their study lent support to the noticing function of output hypothesis. However, there was a crucial limitation to this study. Output group actually received more time on task than the non-output group. This made it hard to claim that the better performance in recall study was facilitated by output activity alone. The treatment procedure in table below shows the difference.

Table1. Treatment procedure in Suzuki et al. (2009)

	Output group	Non-output group
1	Picture description writing (10 min)	
2	Read model story with pictures provided (3 min)	
3	Performs math task (2 min)	
4	Recall model story (7 min)	

(From Suzuki et al., 2009, modified by the author)

These findings on language proficiency and its role in output triggered noticing seemed to be mixed. While the high proficiency group in Hanaoka (2007) did not outperform less proficient group, Suzuki et al (2009) found that language proficiency did have a facilitative role in output triggered noticing and subsequent learning. However, as mentioned above, both studies have their own limitations and besides both studies did not divide their groups on an objective standard (i.e., some sorts of test scores). It will be much preferable to have two groups divided by a same standard.

Surprisingly, no studies have ever examined the relationship between learner's cognitive abilities and noticing in light of the output hypothesis. Future studies are definitely needed to address the relationship between these individual difference factors and noticing triggered by output tasks.

2.5 Summary of the findings in previous studies

In sum, the previous studies have all been trying to address following two fundamental research questions:

(1) Does output promote noticing?

(2) Does output facilitate learning?

Among them some tried to examine the mediating effect of factors such as complexity of linguistic structure (Uggen, 2012), task type (Ghari & Moinzadeh, 2011; Song & Suh, 2008), task demands (Borjigin et al., 2013) and language proficiency (Hanaoka 2007; Suzuki et al., 2009) on output triggered noticing and subsequent learning.

With regard to the first questions, the results were mixed. While some of them were able to come up with supporting or partial supporting evidence (e.g., Song & Suh 2008; Uggen, 2012) some did not.

In terms of acquisition, the findings were also mixed. While Izumi (2002), Uggen (2012) and Borjigin et al, (2013) provided a positive effect of output on learning the target forms, others failed.

2.6 Remaining problems

Although some studies have provided support for this function of the output, majority of them did not. Thus, further investigation is still necessary. Izumi (2002) has claimed that depth and level of noticing is more crucial for learning than amount of noticing. Therefore, in future studies it is necessary to look at factors which are influencing the depth of processing. For example, as Ranta (1998) found, learner's cognitive ability (namely, the language analytical ability) may have a role in noticing and learning of the target forms. Besides, task factors, such as task type difference, task demands may also have their role in mediating the effect of output on noticing and acquisition. In addition, the cognitive demand from the complexity of linguistic structure also needs to be researched.

2.7 Research questions

In an attempt to address some of the above mentioned issues and to provide a better understanding of the effects of output on noticing and acquisition; and the possible role task factors, along with learner factors, which might have roles to play in such processes, the following two major research questions along with the sub-research questions form the basis of this study:

1. Do output tasks promote noticing of the target form?

If so:

(1) Which of the two output tasks better facilitates noticing of form?

(2) Is such effect likely to be mediated by learner factors such as language proficiency and language analytical ability?

2. Do output tasks facilitate acquisition of the target form?

If so:

(1) Which of the two output tasks better facilitates acquisition of the form?

(2) Is such effect likely to be mediated by learner factors such as language proficiency and language analytical ability?

Chapter 3 Methodology

This chapter details the research design and procedure of the study. An overview of the experimental design is provided first, which is then followed by a description of the participants who volunteered to take part in the study. Next, the treatment procedures and materials are described in detail. Finally, scoring criteria are explained.

3.1 Overview of the experimental design

The study followed a pre-test, treatment, post-test and delayed-post-test design and included three groups: a cloze group, an editing group and a non-output group. The between-group factor is task difference (cloze task; editing task; non-output task).

The within group factor is time (pretest and posttests). The dependant variables are noticing scores and acquisition measures (test scores). The entire procedure, including testing and treatment sessions, took approximately four hours and spanned about four weeks. A delayed posttest was administered approximately two weeks after the immediate posttest.

3.2 Participants

The participants were 88 (60 females and 28 males) college students in Inner Mongolia Autonomous Region, China. After deleting some incomplete data which was due to some of the participants' absence from one or more sessions during the experimental procedure, 79 (59 females and 20 males) of them were able to contribute to the final data analysis. Table 3.1 summarizes the groups and the numbers of participants in each group.

Table 3.1 Numbers of participants in each group

Groups	Non-output	Cloze	Editing
n	23 (5)	28 (6)	28 (9)

Note. Numbers in parenthesis stand for the number of male participants

All of them were freshmen from two different classes of a national university and majoring in economics related subjects such as international economics and trade, financial management, and labor and social security. They were all of Mongolian ethnicity, and learning English as a compulsory course. In Inner Mongolia Autonomous region, there are two types of schools from primary level to tertiary level: Chinese schools, where all subjects are taught in Chinese language; and Mongolian schools, where all subjects, except Chinese, are taught in Mongolian language. The participants in this study have all been educated in the Mongolian schools where they have to learn three languages, Mongolian, Chinese and English almost at the same time from the elementary school (i.e., Mongolian and Chinese from the first grade and English form third grade).

According to the university curriculum, during the first year, the English class would be taught twice a week, every time for 90 minutes, making a total of 180 minutes of class time per week. And once they become sophomores, the English class would be reduced to once a week which makes class time only 90 minutes per week. All of them have been learning English for over six years, with majority of them (82%) for over nine years. In Inner Mongolia Autonomous Region, English is introduced in the elementary school curriculum in most of the Mongolian schools in the most of the area, but there are some schools that teach it from junior high level. That is why 15 out of 88 students who have participated in the present study have studied English for six years not nine as the rest do. However, since the number was small and the linguistic form targeted in the study (i.e., passive voice) would not be introduced until junior high school curriculum, it was considered appropriate to overlook this difference of the length of English learning.

3.3 Experimental Procedure

The whole experimental procedures were carried out during the participants' regular class time. On Monday of the first week of the experiment, after being introduced to the class by the teacher who was teaching their English, the researcher made a brief self-introduction, and then explained the aim of the experiment, in which she informed the class that the experiment was to be done in order to collect data for the doctoral dissertation of the researcher, so they did not have to worry

about the outcome of the tests and tasks, but should try to do it as carefully as possible. Then, the overall schedule of the experiment was explained, in which the class was told that the researcher would come to visit them during their regular class time three times over two weeks, and then on the forth week, with the absence of researcher, there would still be another session to be carried out under the supervision of their English teacher. This final session referred to the delayed posttests, which were carried out by the English teacher of the class without the presence of the researcher. The English teacher of the class had been present during all of the experimental procedure to help the researcher to maintain class order and after each experimental session, carried out her regular teaching. Finally, the researcher asked them to sign the consent form as required by the Tohoku University, Graduate School of International Studies. Then the participants were involved in filling out the back ground information questionnaire. This took about 15 minutes. Language analytical ability test was followed, which took another 15 minutes. After these procedures had been completed, the researcher withdrew from the classroom and the participants went on with their regular class.

The following day, the pretests were administered, which took approximately one hour of the class time and after that the participants went on to have their regular class as well. Since there were two different types of tests (i.e., multiple-choice test and picture-cued written output test) in the study to measure the acquisition of target linguistic form (i.e., simple past tense in English passive voice), each test took almost 30 minutes to be completed. The participants were allowed 10 minutes break between the two tests.

The following Monday, the treatment session and the immediate posttests were carried out. First, the researcher explained the overall procedure of the experimental session that there would be two big session, one was for treatment session and the other one test session. The participants were also told the approximate time necessary for each session. Then the treatment packages were distributed randomly. The treatment packages were arranged in this way that if the first package was for non-output task, then next for cloze task, and then editing task package was followed. In order to make sure it could be evenly distributed, the researcher herself handed out the package one by one alone. In this way, the class was randomly divided into three groups: the non-output group refers to the students who received true or false task packages; the cloze reconstruction group refers to those who received cloze reconstruction task packages; and the editing

reconstruction group, refers to those who received editing reconstruction task package.

After receiving the package, the participants were told not to open it until they received instructions from the researcher. The treatment procedures were carried out according to the following instructions. The actual manual employed during the experiment was a Chinese version (see appendix 3), which was the exact translation of the English version below.

(1). The results of the tasks and the following tests have nothing to do with your overall class performance, so just try your best to do the tasks and do not worry about the results.

(2). Please make sure that you have written your student ID number on the cover page. You do not have to write your name.

(3). Now please open the cover page, and turn to the first page. Read the instruction carefully and then read the passage. The time limitation is 3 minutes. You are not allowed to turn to the next page before receiving instruction.

1 minute left.

30 seconds left.

(4). Time's up. Now turn to the next page and do the task according to the instruction. The time limitation is 5 minutes. You are not allowed to turn back to previous page nor turn to the next page.

1 minute left.

30 seconds left.

(5). Time's up. Please turn to the next page and read the passage again. The time limitation is 3 minutes. Again, you are not allowed to turn back to previous page nor turn to the next page.

1 minute left.

30 seconds left.

(6). Time's up. Now please turn to the next page, which will be the last page of the package, and

do the task according to the instruction. You are not allowed to turn back to the previous page.

1 minute left.

30 seconds left.

(7). Time's up. Thank you for your cooperation. Please pass the packages from the back row to the front.

(8). Please have a 10 minute break.

After the break, the immediate posttest session was followed, which took approximately one hour. There was another 10 minute break between two types of tests.

The following week, which was the third week, nothing took place. Then on Monday of the forth week, the delayed posttests were administered without the presence of the researcher. Since the English teacher has been present throughout all previous sessions and was already very familiar with the procedures of the experiment, it was considered appropriate to administer the tests under her supervision. This session was conducted just as the same with the previous two test sessions, with each type of test taking 30 minutes and 10 minute break in between. The whole experimental procedure is summarized in table 3.2 and treatment session is illustrated in figure 3.1.

Table 3.2 Experimental procedure of the study

Fisrt Week		
	Monday	Consent form
		Background information questionnaire
		Language aptitude test
	Tuesday	Pretests
Second Week		
	Monday	Treatment
		Immediate posttest
Forth Week		
	Monday	Dealyed-posttest

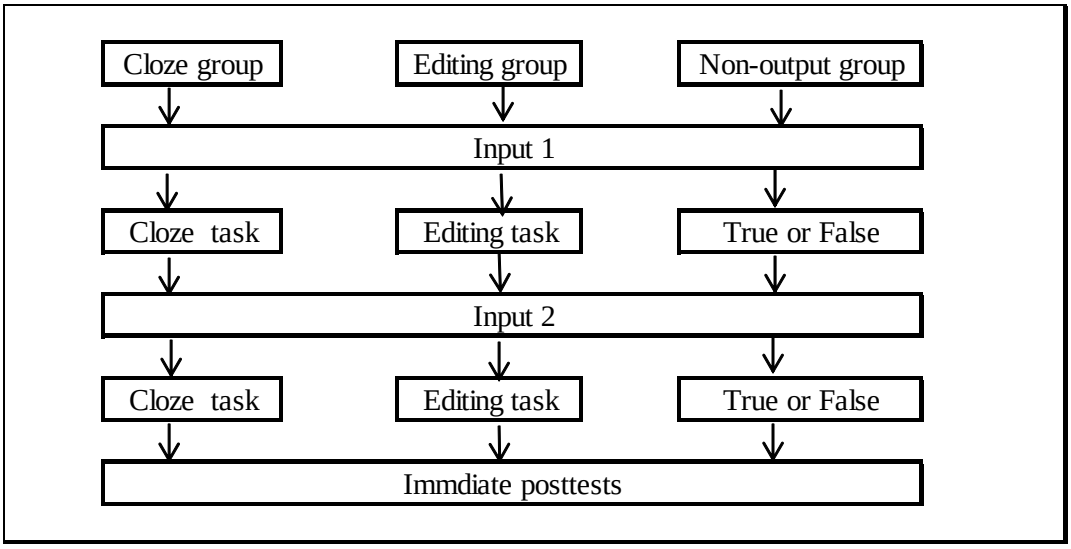


Figure 3.1 Treatment procedure of the study

3.4 Materials

3.4.1 Target structure

The simple past tense of passive voice (e.g., She was employed as a secretary) was targeted in the present study owing to several reasons. First, the passive voice has been considered as one of the most difficult grammatical structures to both teaching and learning in English (Hinkel, 2002).

Secondly, in output hypothesis, it is claimed that pushed output could provide learners with the opportunity to realize the gap between their interlanguage knowledge and the target language, which indicates that learners should already have partial knowledge of the target linguistic form, but not yet perfect. Since the participants in the study were taught passive voice in their junior high school and the scores of the pretests indicate their partial knowledge of it, which makes it an appropriate choice. Finally, findings from Borjigin et al (2013) suggested that it is better to target only one specific form, rather than targeting many. In their study they included various tenses in the passive voice, which is considered one of the reasons to have caused heavy cognitive demand on the participants and thus hinder their acquisition of the passive voice structure. Therefore, the present study focuses on only simple past tense of English passive voice which is structured *be* + past participle of verbs.

3.4.2 Input passage

The input passage along with its picture was chosen from a text book called Side by Side, third edition, book four with simplification in order for participants to comprehend and memorize easily. The simplifications were made in several ways. First, the name of the girl in original text was Amelia, which was shortened in the present study to Amy. Second, all the complex sentences with clauses were simplified except the last sentence.

The modified input passage comprised of 134 words including the title. There were ten sentences in the passage, with every one of them including one passive structure, and two of them containing passive voice indicator *by*.

The picture was provided along with the input passage to help learners to memorize easily. To lessen the burden of the students to recall the input passage when they carry out the required tasks, the picture was also provided in three types of task sheets as well. You can see the input passage and the picture below:

A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

3.4.3 Reconstruction tasks

As has been repeatedly pointed out in previous studies (Hanaoka, 2007; Izumi 2002; Izumi et al, 1999; Muranoi 2007)

reconstruction task is considered one of the most effective techniques to elicit learner output, especially when there is a specific linguistic focus in the study. Through reconstruction, learners will be able to compare their interlanguage (IL) and target language (TL), which may lead to their noticing of the target forms. As Izumi (2002) pointed out "…one advantage of reconstruction task lies in its control over the content and form that learners produce" thus "…maximizing the

equivalence between the learners' output and the target input …" (p. 551).

Besides, it is also considered effective in input-output-input-output treatment design as in the present study hoping to capture the difference between learners' noticing from first to second input, and uptake from first to second output.

3.4.3.1 Cloze reconstruction task

In present study, reconstruction task is operationalized in two different ways. One is in the form of cloze task, or sometimes considered as a fill-in-gap task. The task requires participants to fill in the missing words according to what they have read without allowing them to go back to the input passage, though the picture from input passage was provided as clues to help them recall. This version is developed based on the findings of Borjigin et al (2013) in which they included two different versions of reconstruction tasks, with one of them asking participants to reconstruct the whole passage (i.e., text reconstruction) and the other asking them to reconstruct only the missing parts (i.e., cloze reconstruction) which were the target forms. They found the latter one facilitated the acquisition of the target form while the previous one failed to do so, which led them to argue that "the cognitively less demanding output task enabled participants to focus more attention on target forms and acquire them" (p. 67).

Therefore, the same form of cloze reconstruction task is adopted in this study and designed with 7 items as distracters and 12 on simple past tense in passive form which include the indicator *by*. The task is shown below.

A very exciting year

In January, Amy ( ) as ( ) secretary ( ) the Inter-Tel company.

In March, she ( ) to school ( ) the company to study statistics and information technology.

In April, she ( ) a raise.

Just two months ( ), she ( ) to the position of supervisor of her department.

In August, she ( ) as the "Employee of the Month".

In October, she ( ) another raise.

In November, she ( ) to apply for a position in the company's overseas office ( ) Bangkok, Thailand.

In December, she ( ) the new position ( ) Thailand.

At ( ) month, she ( ) the "Employee of the Year".

She can't believe all the wonderful ( ) that ( ) to her since she ( ) just twelve months ago.

3.4.3.2 Editing reconstruction task

The other form of reconstruction task was carried out as an error correction editing task, also called in many studies error correction or form correction. It requires participants to identify the errors and then based on what they have read try to edit the erroneous part into correct one. Again, as has been done in the cloze task, the participants were told to provide the correct form without referring back to input passage. Error correction is a commonly practiced task in language classrooms as cloze task (Lee, 2007; Storch, 1999), but was considered having an advantage in drawing learners' attention from meaning to form (Lee, 2007; Wong, 2003) since it asks learners to identify and correct the errors.

The numbers of distracters and passive forms are kept the same as the cloze task. The erroneous parts in this task are the parts that are missing in cloze task. The types of error targeting passive forms following the previous studies (Lee, 2007; Spada et al, 2014) were designed in following ways:

(1). absence of be verb;

(2). using present participle instead of past participle of verb after *be* verb;

(3). active voice instead of passive voice;

(4). using original form of the verb instead of past participle after *be* verb;

(5). incorrect form of past participle.

Therefore, among 10 sentences, three of them contain type (1) error, one of them contains type (5)

error, rest of the six are divided equally into error (2) (3) (4). The distributions of the errors are shown below in the task.

A very exciting year

In January, Amy employed as the secretary for the Inter-Tel company. (1)

In March, she sended to school with the company to study statistics and information technology. (5)

In April, she was giving a raise. (2)

Just two months late, she promoted to the position of supervisor of her department. (1)

In August, she chooses as the “Employee of the Month”. (3)

In October, she gave another raise. (3)

In November, she was inviting to apply for a position in the company’s overseas office on Bangkok, Thailand. (2)

In December, she provided the new position of Thailand. (1)

At end of month, she was award the “Employee of the Year”.(4)

She can’t believe all the wonderful thing that has happened to her since she was hire just twelve months ago. (4)

However, compared to the cloze task, which requires participants to reconstruct the whole passive structure (*be* + past participant) the editing task already provided either *be* or past participle, though some of them were provided in incorrect form. Thus, the scoring criteria for the two types of tasks were different.

There are two underlying rationale for these two reconstruction tasks. First, as has been mentioned in chapter 2 that Robinson et al (2012) suggested that in order to examine whether a certain individual difference associate with a particular task character, it would be preferable to keep the task difference at a minimum level. Therefore, in present study the two output tasks are designed to differ from each other in one major way. The Cloze task requires reading, recall and reconstruction steps while editing requires reading, recall, compare and reconstruction. Then one may expect that since the missing parts are so obvious in the cloze reconstruction task, it might be hypothesized that those who were involved in the cloze reconstruction task would underline more

passive related items in second input phase compared to those who were engaged to editing reconstruction. Editing reconstruction, on the other hand, might be hypothesized to be more effective in helping learners process the form better, since it requires a comparison stage, which put in other words, provide learners with opportunity to compare the mismatch between erroneous forms and the correct forms. Secondly, previous studies (Nassaji & Tian 2010; Storch 1997; 2007) also suggested that despite being well-used FFI tasks the effects of cloze and editing tasks have not been thoroughly researched.

3.4.4 True or false task

As for the non-output group, after reading the passage, they were asked to do a true or false task which is a well used reading comprehension task in language classrooms. The sentences provided in the true or false task avoided using passive structures in order not to provide extra input on the target form. The task was shown below:

( ) 1. Amy is a sales woman in a telephone company.
( ) 2. Amy went to school to study foreign languages, because the company asked her to do so.
( ) 3. Amy became the supervisor of her department in July.
( ) 4. The company has given raises to her three times within a year.
( ) 5. Amy wanted to apply for the new position but the company said no.
( ) 6. Amy quitted her job and moved to Bangkok, Thailand to get a new job.
( ) 7. The company awarded Amy the "Employee of the Month" twice.
( ) 8. Amy had a very exciting year, because she had many chances to go abroad to company's overseas offices.
( ) 9. Amy went to Thailand in November.
( ) 10. Since her employment in January, Amy has had many exciting experiences at work.

3.4.5 Testing materials

It has been suggested in previous studies that more than one type of test should be used to provide a more collectively reliable results and accordingly increase chances of coming up with more complete picture of learners' performance (Spada et al, 2014). Therefore, in order to assess the knowledge of passive voice before and after the treatment session, two types of tests were used. These two tests were supposed to tap into two different types of knowledge about English passive voice. A multiple choice test was designed to provide a measure of participants' receptive knowledge (or sometimes called comprehension test) and a picture-cued written output test (or sometimes called production test) was intended to examine the participants' productive knowledge.

3.4.5.1 Multiple choice tests

Multiple choice questions are not only a well used type of test in schools, but also a very common type of exercise in textbooks. Thus the participants are well acquainted with this type of test. Besides, it has following advantages: (a) well suited to test discrete features of grammatical knowledge; (b) easy administration and (c) objective scoring (Purpura, 2004). The multiple choice test used in the present study was based on the one that had been used in the researcher's master thesis but has been modified, since the linguistic target of that study were various tenses of passive voice whereas only simple past tense in passive voice was targeted in this study. However, all the distracters remained the same with the previous version. Altogether there were 10 items on target linguistic structure, that is, simple past tense in English passive voice, and 20 items on various linguistic features as distracters. Every item has four options. Despite the correct option *was/were + past participle of the verb* (e.g., was injured), other three options were designed following previous studies (Lee, 2007; Spada et al 2014):

(a) past form of the given verb (e.g., injured);
(b) had+ past participle of the given verb (e.g., had injured);
(c) was/were + original form of the given verb (e.g., was injure);
(d) was/were + present participle of the given verb (e.g., was injuring).

Since the participants were tested three times as they participated in pretest, immediate posttest and delayed-posttest, two different versions of multiple choice tests were designed. The version one used as pretest was again administered as delayed posttest because there were 4 weeks gap between the two sessions and the period is considered long enough for the participants to forget about the items.

The distracters maintained the same in both versions of tests while the 10 target items were modified in version two, which was used as immediate posttest. Two of the verbs from version one remained the same in version two but the sentences were different. The sentences were:

From version one (pretest and delayed posttest)

21. I ____ to the party, but I went anyway.

A. did not invite B. was not inviting C. was not invited D. did not invited

24. The new classroom building _____last month.

A. completed B. was completing C. had completed D. was completed

From version two (immediate posttest)

19. I_____to the house warming party of my new neighbor, but I went anyway.

A. did not invite B. was not inviting C. was not invited D. had not invited

28. The document _____ and everything was ready for the meeting.

A. completed B. was completing C. has completed D. was completed

Apart from these two verbs, all the rest were different from each other but the numbers of regular and irregular verbs were kept equal. There were seven regular verbs and three irregular verbs in each version. The design of the four options for each items remained the same (see appendix 1 for

the testing materials).

3.4.5.2 Picture-cued written output test

This test was originally developed by Spada et al (2014) as an oral production test aiming to measure the intuitive (i.e., implicit) knowledge of the participants. In their study, they also included two language measures: a written error correction task (ECT) and a picture-cued oral production task (OPT). The ECT is considered to be a measure of controlled analyzed knowledge of grammar (i.e., explicit knowledge) and the OPT as one in which students spontaneously draw on their intuitive (i.e., implicit) knowledge of grammar in communicative interaction. However, in the present study with the exact same pictures written form of production was required for the participants rather than oral form, since the research questions addressed were totally different and even the aim of the test itself was also different, with Spada et al (2014) measuring implicit knowledge of students while the present study interested in measuring productive knowledge.

The same version of the picture-cued output test was used throughout all testing sessions, assuming that participants may not be able to memorize the pictures under the time limitation and pressure of producing the sentences. Besides, there was enough length of time interval in between test sessions with one week interval between pretest and immediate posttest and another two weeks between posttest and delayed posttest. This picture-cued written output test was a storytelling task about a package that was lost in the mail. Sequences of pictures were arranged in a complete story line beginning from a picture aiming to elicit “The package was taken to the post office”, “The package was weighed” to undergo some process in which the package has been mistakenly delivered to different places then returned to the right place in the end. Sentences such as “The package was returned to the post office” and “The package was recycled” were supposed to be produced by the participants. There were altogether 16 slides in the test with four of them were designed to give a overall introduction to the whole story and ten of them were intended to elicit simple past tense in passive voice and two of them active sentences. The slides were delivered via Power Point on a big screen to the whole class one by one. The participants were asked to write down one sentence describing each picture within 1 minute of time, using the verb provided on the top of every slide. The actual instruction during the test was in Mongolian and

occasionally with the help of Chinese words so that the researcher could avoid using any passive voice to give the overall introduction of the story. The English translation of the manual was originally borrowed from Spada et al' (2014) study with a little modification in this study which was detailed below.

(1). (Start the PowerPoint slides. Go to the first slide (what happened to the package).)
Researcher: "I'm going to show you some pictures that tell a story about a package that got lost (this particular expression was translated into the Chinese word 'diu' which indicates the similar meaning as got lost but not passive) in the mail."

(2). (Show slide 2 with Anna and her mother on it)
Researcher: "The name of the story is {what happened to the package?}
In this story, Anna has moved to Canada from Mexico. Now she lives in Toronto. By the way, do you know the fact that actually there are two cities named Toronto in the world? One is in the USA, and the other one is in Canada. Anna is in Toronto of Canada. The rest of her family still lives in Mexico. Her mother is sending Anna a package of her favorite things. Unfortunately, this package got lost in the mail."

(3. (Go to slide 3 and point at it.)
Researcher: "These pictures tell the story about how Anna's package got lost in the mail. I'm showing the pictures together so you can understand the story and then I'll show the pictures to you one at a time and you can write down the sentence describing each picture one after one. " "Please take a moment to look at these pictures – they tell the story about what happened to the package from the beginning to the end."

(4). (Wait for learner to do this)
Researcher: "Do you have any questions about the pictures? For example, you cannot see some pictures clearly or have difficulty with some of the given words, etc." (Point to the "bring" picture. Answer any questions.)

Researcher: “I want you to write down the story using the verbs at the top of the pictures (point to verbs). Do you understand the verbs that are on the pictures?” (Point to “weigh”) Do you understand the verb weigh? (Answer any questions about the meaning of the verbs.)

(5). (Continue with the slide show)

Researcher: “Now I’m going to show you one picture at a time. Please write down the sentence describing the picture one by one. OK? Remember, use the verbs provided in each picture. You have 1 minute to write”

“So what happened to the package?”

“30 seconds left.”

“Time’s up. Now let’s move on to the next picture. ”

“30 seconds left.”

The last two steps were repeated until all the pictures were shown.

3.4.5.3 Language analytical ability test and language proficiency test

Language analytical ability test in this study was originally developed by Ottó aiming to measure the language analytical ability of the learners. The test has been frequently used in the previous studies (see Schmitt et al 2003; Sheen 2007). There are 14 items in the test with a multiple-choice format (see appendix 1.1). In the first part, the words and sentences in the artificial language and their English translations were provided. In order to make sure that the participants understand what they were supposed to do, the researcher demonstrated how to do such kind of test with the example question given in the first part. The researcher also confirmed if the participants were familiar with all the English words and sentences in the test because English was not their first language. After all these procedures, the participants were asked to carry out the 14 items given in

the second part of the test. One point was given to each correct choice and that made the total score 14 points.

Language proficiency test scores of the participants in the study were actually their end-of semester test scores because these scores demonstrated a strong correlation with their college entrance test scores (r (74) = .57, p < .001). The end-of-semester test consists of four parts: listening (15 points); reading (40 points); grammar (30 points); and translation (15 points).

3.5 Scoring criteria and data analysis

The data collected during the tasks and testing sections made up the corpus of analysis of the study. The description of each data set and how it was analyzed are explained below.

3.5.1 Scoring of the tests

3.5.1.1 Multiple choice tests

Only the 10 target items out of each test (pre, post and delayed posttest) were scored. One point for each correct choice made the total points to 10.

3.5.1.2 Picture-cued written output tests

The criteria for scoring picture-cued writing were adopted from Spada et al (2014) with some modification. The modified parts were shown in italics. The first modification was made in (2), where originally *tense* was not included. However, the test was originally used as an oral production test and there was communication between researcher and the subject during the test. The researcher in the original study provided feedback which constantly included past tense so the subject would more likely to produce sentences in past tense while in the present study no such support or feedback was provided during the test, it was considered appropriate to overlook the tense mistake.

The second modification was made in (4), where a new criterion *e* was added. It was because

the teacher who was teaching English to the participants of the present study pointed out that it was a common error made by her student when they were expected to produce an active sentence. The students tend to produce sentence including both *be* and actual verb when they intended to make an active sentence and she provided some proof for it by showing some extracts from the participants' ordinary class assignments:

Examples:

Student A: College students are worry about how to get a job.

Student B: I will be get good grades and win the scholarship.

Student C: Competence is become more and more important.

Student D: college students are face as many kinds of pressure.

In all of these sentences the students were obviously trying to produce some active sentences, however, as has been pointed out by the teacher, they used *be* and actual verb together. It was also added by the teacher that no matter how frequently she provided them the feedback they made the same mistakes again and again which indicates that the students have imperfect internalization of the use of verbs when producing active sentences. Therefore, the combination of *be* and original form of the actual verb was considered as an active voice sentence in the present study. The specific criteria were detailed below.

(1) If the learner produced a correct passive sentence, 3 points were given.

(2) If the learner produced a passive sentence with errors in agreement, *tense* and/or past participle (e.g., "The package is/was sended to Canada."), 2 points were given.

(3) If an active instead of a passive sentence was produced (e.g., "They put the package on the truck."), 1 point was given.

(4) No points were given in the following instances:

a. if an incorrect active sentence was produced (e.g., “They putted the package on the truck.”),

b. if the present participle was used instead of the past participle (e.g., “The package was sending to Canada.”),

c. if the present perfect auxiliary *have* was used instead of the auxiliary *be* (e.g., “The package have deliver.”),

d. and/or if illogical sentences were produced (e.g., “The package is they put US post office.”), 0 points were given.

e. *if the original form of a verb was used after the auxiliary be (e.g., “The package is/was/be deliver.”)*

3.5.2 Scoring of tasks

3.5.2.1 Scoring of underlining

(1) If a complete passive form that is, *be* + past participle, was underlined 1 point was given.

Example:

Amy <u>was employed</u> as a secretary by the Inter-Tel company.

(2) If the passive voice indicator *by* was underlined 1 point was given.

Example:

Amy was employed as a secretary <u>by</u> the Inter-Tel company.

(3) If either auxiliary *be* or a past participle was underlined, 0.5 point was given.

Example:

Amy was employed as a secretary by the Inter-Tel company.

Amy was employed as a secretary by the Inter-Tel company.

Table 3.3 Linguistic forms targeted in both cloze and editing reconstruction tasks

	Cloze	Total Points(20)	Editing
1	was employed; a; by	3	was employed; a; by
2	was sent; by	2	was sent; by
3	was given	1	was given
4	later; was promoted	2	later; was promoted
5	was chosen	1	was chosen
6	was given	1	was given
7	was asked; in	2	was asked; in
8	was provided; in	2	was provided; in
9	the end of the; was awarded	3	the; the; was awarded
10	things; have happened; was hired	3	things; have happened; was hired

As been shown in the table above, 12 out of 20 points were on passive form related items, which include the auxiliary *be*, past participle of a verb and the indicator *by*. Since both cloze and editing reconstruction tasks were designed to target same linguistic forms, the points were the same. However, specific operation of the scoring was different because of the difference of the task type.

3.5.2.2 Scoring of cloze task sheet

The evaluation of the task sheets were done in three different categories:

(1) total scores of correctly reconstructed items which include both passive structure and distracter items;

(2) passive structure scores only which include correctly reconstructed complete passive

structures (*be* + past participle) or incomplete with either the *be* or past participle and indicator *by*.

In addition, the numbers of attempts were also provided which include both complete and incomplete passive structures, and incorrect form of past participants.

Scoring example of cloze reconstruction task:

Total score: 5points
Passive structure score: 5 points
Attempts: 5 times

In January, Amy (was employed) as (a) secretary (by) the Inter-Tel company. (1+1+1 points)
In March, she (sent) to school (by) the company to study statistics and information technology. (0.5+1 points)
In April, she (was) a raise. (0.5 points)

3.5.2.3 Scoring of editing task sheet

The evaluation of the editing reconstruction task sheet was also done in three different categories as cloze reconstruction task:

(1) total scores of correctly reconstructed items which include both passive structures and distracter items;

(2) scores of passive structure only which comprised of the following points:

(a) insertion of auxiliary *be*;
(b) reconstruction of correct past participle of the verb;
(c) successful conversion of active voice into passive voice;

(d) reconstruction of indicator *by*; and

(3) numbers of attempts which include any kinds of attempts in reconstruction of passive structures, correct or incorrect.

Scoring example of editing reconstruction:

Total score: 3 points

Passive structure score: 2 points

Attempts: 5 times

In January, Amy ∧ employed as the secretary for the Inter-Tel company. (1 point)
was

In March, she sended to school with the company to study statistics and information
send

technology. (0 point, since the past participle form of send was not correct)

In April, she was giving a raise. (1 point)
given

Just two months late, she promoted to the position of supervisor of her department. (1 point)
later

In August, she ∧ chooses as the “Employee of the Month”. (0 point, unsuccessful conversion of
was

active sentence into passive)

In October, she gave another raise. (0 point, unsuccessful conversion of active sentence into
given

passive)

The difference in second scoring criterion in two types of task made it inappropriate to compare the total points and the passive structure points directly. However, the way of measuring attempts remained the same which enabled the comparison of the numbers of attempts possible for two output groups in order to provide a different perspective to examine noticing of the target form other than only looking at their underlining scores.

Chapter 4 Results

This chapter presents the results from the study. The content is arranged in the order of the research questions, that is, noticing issue first and acquisition next. Thus, the first part presents those results of underlining scores and task sheets performance which address the noticing issue. Second part reports the results of tests which address acquisition issue. The descriptive data are presented first and inferential statistical results follow. Finally, the relationship between noticing and acquisition is presented. The discussion of the results is presented in chapter 5.

4.1 Noticing Results

Noticing of the target linguistic form (i.e., simple past tense in passive voice) is measured through two different perspectives. First, it is measured through comparing the scores of first underlining which is completed before the tasks and the second time underlining which is done right after the tasks of all three groups. Then, it is also measured by comparing the improvement of attempt scores from first to second time task sheets of the two output groups.

4.1.1 Results of underlining

The following table summarizes the mean scores along with the standard deviations of the first and second time underlining scores of each group, and the figure illustrates the change of the underlining scores of each group from first to second time.

Table 4.1 Mean and SD of first and second time underlining (noticing) scores of each group

	first time	second time
	Mean (SD)	Mean (SD)
Non-output (n=23)	3.26 (2.44)	2.78 (2.13)
Cloze (n=28)	2.89 (2.25)	5.04 (3.80)
Editing (n=28)	2.36 (1.91)	4.36 (1.95)

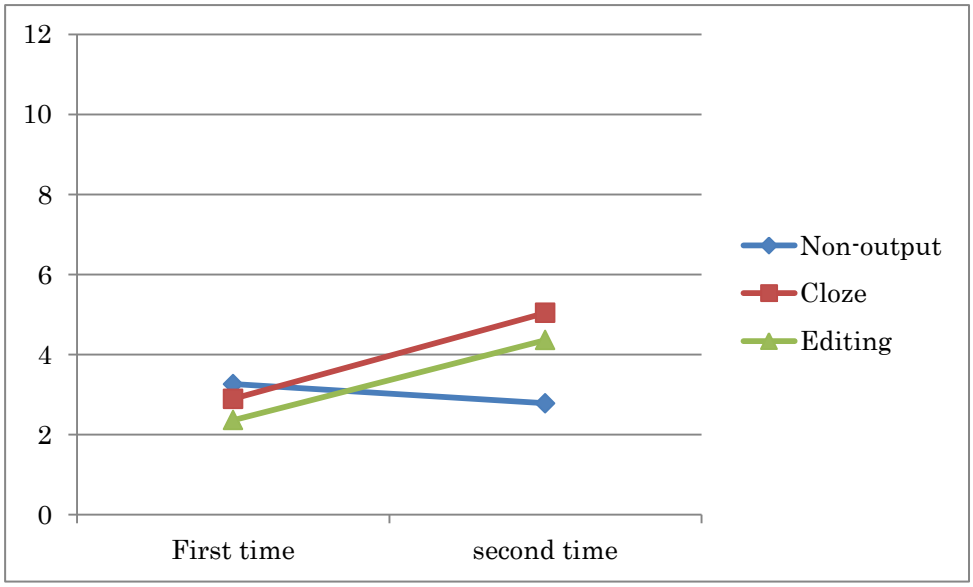


Figure 4.1 Mean scores of first and second time underlining of three groups

As the table shows there is an increase between first and second time underlining scores for both output groups. Non-output group, on the other hand, has failed to achieve the same. This indicates that the participants who were involved in two output tasks underlined more items related to passive voice in the subsequent input activity. The results were submitted to one-way ANOVA analysis to examine whether there was any significant difference between the groups.

The results revealed that there was no significant difference between scores on first underlining (F (2, 76) = 1.10, p = .34) which means before any of the tasks, the noticing scores were same. However, significance was found on second underlining scores (F (2, 76) = 4.25, p = .02) which took place after each group carried out the required task respectively. This suggests that those who were engaged in doing output tasks managed to outperform those who were engaged in

non-output task, in this circumstance, a true or false task.

Furthermore, a post hoc comparison using the Fisher LSD test revealed that there was significant difference between non-output group and cloze group ($p<.01$), non-output group and editing group ($p<.05$), while no significant difference was found between cloze and editing group. LSD was chosen because it was suggested to be the most powerful post hoc, especially when there were only three groups in comparison (Larson-Hall 2010). Though both output groups improved on their second underlining scores, and the results of the paired-samples *t* test showed that their improvement from first to second time underlining scores were significant (t (27) = -2.99, p <.01 for cloze group; t (27) = -4.67, p <.01 for editing group), the between group comparison through one-way ANOVA analysis revealed no significant difference (F (1, 54) = 0.71, p = .40).

Therefore, though output groups outperformed non-output group on triggering noticing, which output task is more effective in doing so was not testified. Thus it can be concluded from these findings that output tasks in this study triggered the noticing of the target form compared to the non-output task and consequently lent support to the noticing function of output hypothesis along with many previous studies (Song & Suh 2008; Uggen 2012).

However, there was no evidence found on the differential effect of task type. To further explore this issue, the numbers of attempts participants produced on the output task sheet were compared, which is shown in following section.

4.1.2 Number of attempts in task sheets of two output groups

The table below summarizes the mean scores along with standard deviation of numbers of attempts in first and second output for the two output groups and the following figure illustrates the increase of the numbers from first to second output task of each group.

Table 4.2 Mean and SD of numbers of attempts of two output groups

	first time	second time
	Mean (SD)	Mean (SD)
Cloze (n=28)	3.00 (2.45)	7.14 (3.08)
Editing (n=28)	5.00 (2.67)	7.61 (2.96)

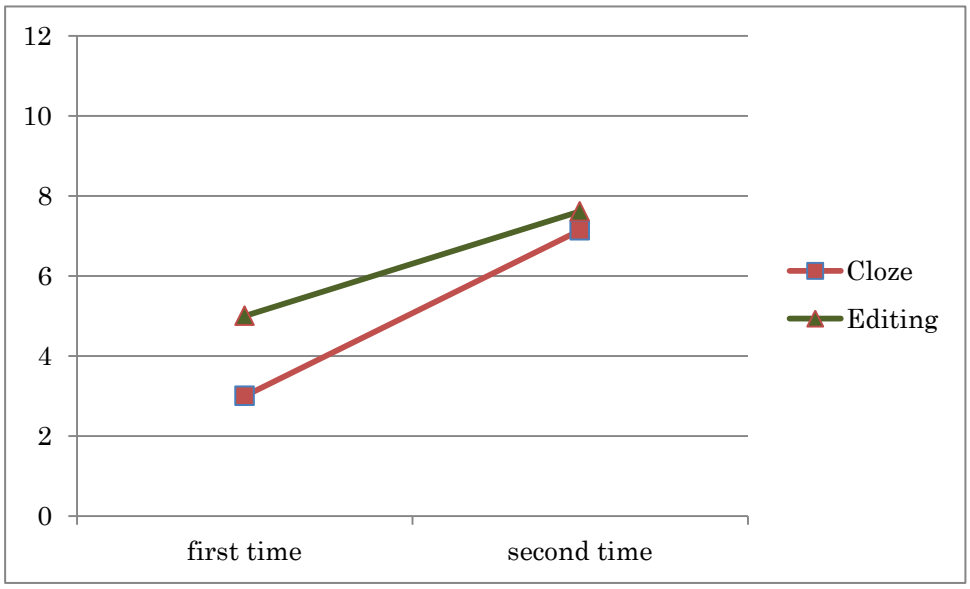


Figure 4.2 Mean numbers of attempts of two output groups

As the table shows, the editing group outperformed the cloze group in the numbers of attempts manifested during their first output. Though the editing group also showed more numbers of attempts on second output than the cloze group, the numbers were very close. However, both groups managed to manifest an increase in the numbers of attempts from first to second output.

The results of the one-way ANOVA revealed that there was a significant difference between numbers of attempts in first output (F (1, 54) = 8.54, p =.005). This indicates that the editing task managed to direct participants' attention to the target forms more than the cloze task did. Therefore, at least, it shows an advantage of editing task over cloze task on directing learners' attention to target form.

However, no significant difference was found between the numbers of attempts during their second output (F (1, 54) =.33, p = .57). A paired-samples t test comparison was conducted to see

whether there was any statistically significant difference between the numbers of first and second time attempts for each group. The results revealed that the increase in numbers were significant for both groups (t (27) = -8.86, p <.001 for cloze group; t (27) = -6.94, p <.001 for editing group). This increase in numbers indicates that both output tasks managed to draw learners' attention to target form though it is unclear which task is more beneficial since though editing task appeared to draw more attention at first output, that was just after the first input phase, it did not show similar advantage in the second output, which was completed immediately after the second input phase.

The reason attributed to this might be the different characters of the tasks. Editing task was designed in an error correction form which would direct learner attention to the erroneous parts regardless of how much engaged the learners in the previous input activity, while the cloze task which requires learners to fill in the missing words or phrases was more closely related to the previous input activity. This also explains why in second output the cloze group managed to improve numbers of attempts significantly and editing task, despite the advantage in first output failed to outperform cloze group. The same can be observed from the figure, the increase in numbers of attempts from first to second time for cloze group is sharper than that of editing group.

Apart from looking into the underlining scores of all groups or the number of attempts made by two output groups, the gain in the respective task of each group should also be examined in order to provide a further evidence to demonstrate the facilitating effect of noticing on the task performance.

4.2 Task performance results

Task performance of each group is presented in this section in order to explore the possible relationship between task performance and acquisition of the target form. The table below summarizes the mean scores along with standard deviation of task performance of all the groups. The figure that follows illustrates the increase in mean scores from first to second task performance.

Table 4.3 Means and SD of first and second time task performance of three groups

	first time	second time
	Mean (SD)	Mean (SD)
Non-output (n=23)	5.48 (1.24)	5.78 (1.24)
Cloze (n=28)	1.54 (2.01)	4.79 (2.82)
Editing (n=28)	2.36 (1.87)	5.36 (3.02)

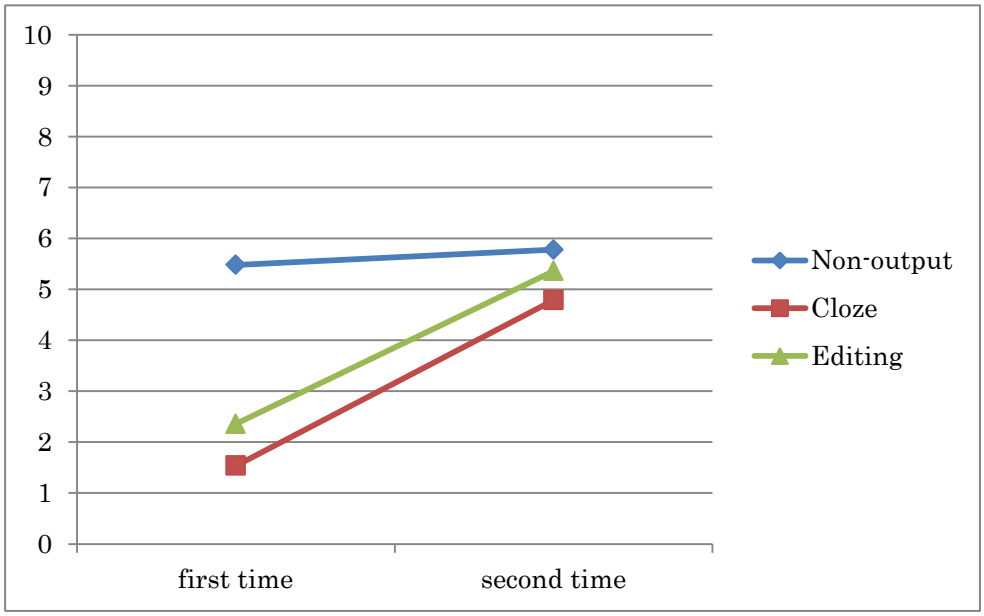


Figure 4.3 Means scores of first and second time task performance of three groups

It can be interpreted from the table that both output groups were able to improve from their first time task performance to the second time. The non-output group, on the other hand, was not able to improve which may attribute to their already high scores on the first performance. Since the different task was scored on different criterion, it is not possible to conduct between group comparisons.

Within group statistical comparison was made to see whether each group managed to achieve significant improvement, and whether this significance in improvement may have effect on their acquisition of the target form.

The results of a paired-samples *t* test revealed that while the improvement of non-output group

was not significant (t (22) = -1.27, p=0.29), that of the two output groups were statistically significant (t (27) = -7.49, p < .01 for cloze group; t (27) = -7.22, p < .01 for editing group). Whether this improvement in task performance for the two output group successfully leads to improvement in acquisition of the target form is discussed in next chapter together with their test results.

4.3 Test results

The acquisition of the target form is addressed through comparing the results of tests administered before (pretest) and after (immediate posttest and delayed posttest) the treatment section. Statistical analyses are conducted to see whether there are any significant improvements from the pretest to posttests. There were two types of tests used in the present study to provide evidence from two different aspect of knowledge of the target form: receptive knowledge, which is measured by multiple-choice tests; and productive knowledge, which was measure by the picture-cued written output tests.

4.3.1 Multiple choice tests results

The table below summarizes the mean scores along with standard deviations (SDs) of pretest, posttest and delayed-posttest of each group. The figure illustrates the increase in mean scores from pretest to posttests.

Table 4.4 Means and SDs of multiple choice tests scores of three groups

	pretest	posttest	delayed posttest
	Mean (SD)	Mean (SD)	Mean (SD)
Non-output (n=23)	4.65 (2.27)	4.78 (2.04)	5.39 (2.52)
Cloze (n=28)	4.04 (2.29)	4.25 (2.55)	4.57 (2.17)
Editing (n=28)	5.21 (1.95)	5.43 (2.28)	5.14 (2.10)

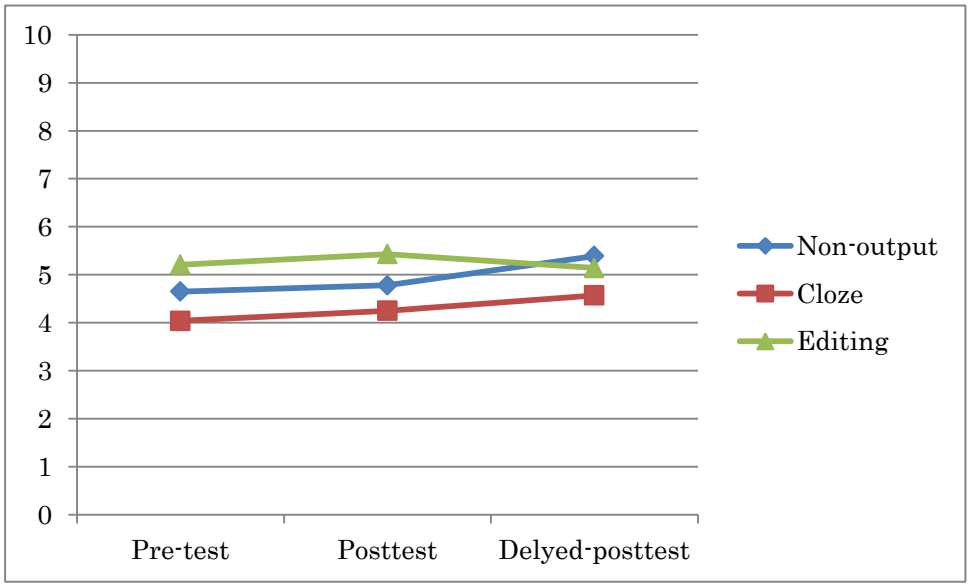


Figure 4.4 Mean scores of multiple-choice tests of three groups

It can be observed from the table that the difference of mean scores between groups do not appear to be significantly different from one another. The scores were submitted to the one-way ANOVA with group (i.e., treatment difference) as between-subjects a factor.

Results revealed that there was no significant group difference in any test (F (2, 76) = 2.07, p =.13 for pretest; F (2, 76) = 1.82, p = .17 for posttest; and F (2, 76) = 0.91, p = .41 for delayed posttest), though the post hoc comparison by Fisher LSD showed that in the pretest stage, there was a significant difference between mean scores of the cloze group and the editing group ($p< .05$). However, no other significant difference between groups was found in post hoc comparison as well.

The scores of each group were then submitted again to one-way repeated-measures ANOVA with time (before and after treatment) as a factor to examine whether there was any significant improvement between pretest and posttests scores within groups.

The results showed that there was no significant increase in scores from pretest to two posttests for all groups. Specifically, for the non-output group, the significant level found in statistical analyses were p = .78 between pretest to posttest, p = .26 between posttest to delayed posttest and p = .10 between pretest to delayed-posttest; those for the cloze group were p = .60 between pretest to posttest, p = .44 between posttest to delayed posttest and p = .17 between pretest to

delayed-posttest; and for the editing group, p = .67 between pretest to posttest, p = .48 between posttest to delayed posttest and p = .87 between pretest to delayed-posttest. The following table summarizes the statistical findings from within group comparison.

Table 4.5 Summary of within group comparison of multiple choice test scores

	pretest-posttest	posttest- delayed-posttest	pretest- delayed-posttest
Non-output	Not significant	Not significant	Not significant
Cloze	Not significant	Not significant	Not significant
Editing	Not significant	Not significant	Not significant

The above results based on the performance on multiple-choice test indicate that output tasks were not able to show a facilitating effect on acquisition of passive form with regard to the receptive knowledge. First, between groups comparisons suggests both output groups failed to outperform non-output group in either posttest or delayed posttest which indicate no beneficial effect of output tasks over non-output tasks. Within the group comparison none of the three groups managed to show significant improvement from pretest to two posttests, although the non-output group displayed a trend toward significance level in their delayed posttest performance. Yet, this improvement was not enough for the non-output group to outperform the two output groups since no significant difference was found between the delayed-posttest scores of the three groups.

Thus, it can be inferred that though output tasks successfully triggered noticing, they failed to show facilitating effect on acquisition of the target form, from the perspective of receptive knowledge. This finding also accords with the previous findings (e.g., Izumi & Bigelow 2000; Song & Suh 2008) that although output does trigger noticing, it repeatedly fails to provide similar evidence when it comes to acquisition issue.

One of the reasons attributed to this may have been as Izumi (2002) has pointed out that it is the depth of noticing that facilitates acquisition not the amount of noticing. Another reason responsible for this finding may be the skill specificity of learning suggested by skill acquisition theory. Dekeyser (1997) found that subjects involved in production practice outperformed those

who were involved in comprehension practice on production test. Thus it is necessary to look into the performance on written output test of the participants in present study to explore whether output facilitates acquisition of the target form.

4.3.2 picture-cued written output test results

Table 4.5 summarizes the mean scores along with standard deviation of pretest, posttest and delayed-posttest of each group. Figure 4.5 illustrates the increase in mean scores from pretest to posttests.

Table 4.6 Means and SDs of written output test scores of three groups

	pretest	posttest	delayed posttest
	Mean (SD)	Mean (SD)	Mean (SD)
Non-output (n=23)	3.35 (5.18)	4.78 (7.88)	7.00 (7.57)
Cloze (n=28)	3.04 (5.14)	5.50 (7.44)	5.46 (8.38)
Editing (n=28)	3.43 (5.57)	4.96 (7.63)	5.50 (8.33)

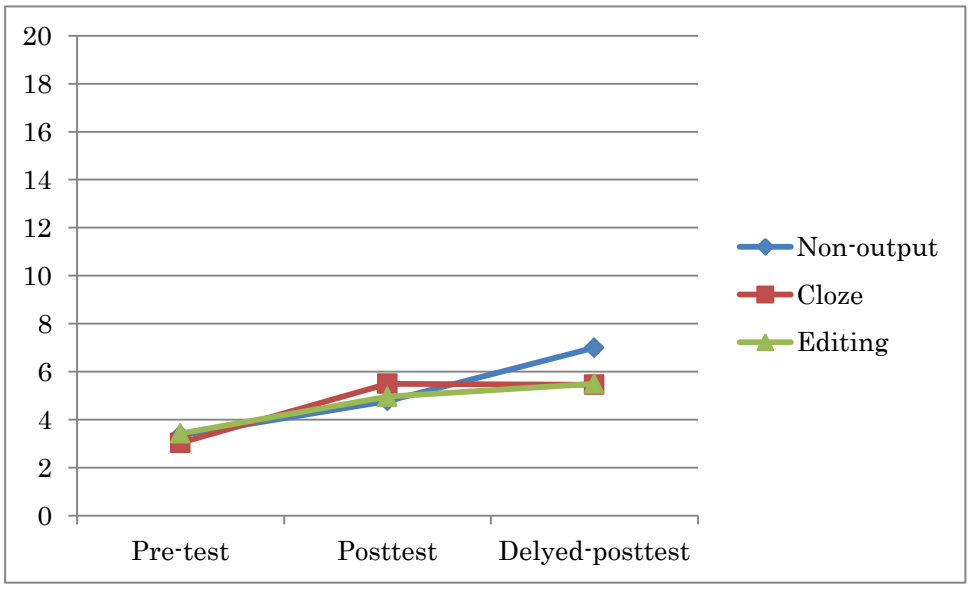


Figure 4.5 Mean scores of written output test of three groups

As the table shows, all groups managed to extend their scores in posttest stage, and the non-output and editing group further increased their scores in delayed-posttest stage as well. Between groups comparison of the scores appeared to be the same at the pretest stage but the cloze group scored more than the other two groups at the posttest stage and at the delayed-posttest stage non-output group scored more than two output groups.

In an attempt to examine whether these differences are statistically significant or not, scores were submitted to one-way ANOVA first with groups (i.e., treatment difference) as within-subjects factor. The results revealed that no significant difference was found between groups at any stage of the tests (F (2, 76) = 0.05, p = .95 for pretest; F (2, 76) =0.06, p = .94 for posttest; and F (2, 74) = 0.28, p = .75 for delayed-posttest respectively). The post hoc comparisons also failed to show any significant difference between any tests and any groups.

The scores then submitted to one-way repeated-measures ANOVA to explore the within group difference between three tests to see whether any one of those groups managed to demonstrate significant difference from their pretest to two posttests. For the non-output group, the difference between pretest and posttest was not significant (p = .11), however, the difference between posttest and delayed-posttest (p = .02), pretest and delayed-posttest scores (p = .004) were both significant.

The results for the cloze group found significant difference between pretest and posttest (p = .003), and pretest to delayed-posttest (p = .009), though not between posttest to delayed-posttest (p = .96). The editing group, on the other hand, did not demonstrate any significant difference between pretest to posttest (p = .07), pretest to posttest (p = .58) and pretest to delayed posttest (p = .12). The cloze group managed to make improvement in posttest, which took place immediately after the treatment, and compare to the scores in pretest, in delayed-posttest stage it also made a significant improvement.

The editing group, the other output group, was unable to make significant improvement over time, both in posttest and delayed posttest. Meanwhile, the non-output group though was not able to make improvement in posttest, was able to make significant improvement from posttest to delayed-posttest and thus pretest to delayed-posttest.

Table 4.7 Summary of within group comparison of written output test scores

	pretest-posttest	posttest- delayed-posttest	pretest- delayed-posttest
Non-output	Not significant	Significant	Significant
Cloze	Significant	Not significant	Significant
Editing	P=.07	Not significant	Not significant

The above results indicate that the output groups failed to outperform the non-output group in written output test either. Although some significant improvements were found in within group comparison, none of those improvements appeared to be significant with regard to the between group comparison. In other words, although the non-output group and cloze group managed to make improvement in posttests, the improvement was still not enough to outperform the editing group.

Therefore, it is necessary to analyze it from different perspectives to see whether any evidence can be found to support the second research question. In search for such a perspective, the attempts made by the participants during their written output test were considered to be an appropriate way since it provides evidence about how many times the participants tried to produce target linguistic forms (i.e., passive form) in order to complete the test.

The following table summarizes the mean numbers of attempts participants made along with SDs on three tests and figure that followed illustrates the increase in the numbers.

Table 4.8 Means and SDs of numbers of attempts in three written output tests

	pretest	posttest	delayed posttest
	Mean (SD)	Mean (SD)	Mean (SD)
Non-output (n=23)	2.09 (2.21)	2.30 (3.02)	3.17 (2.96)
Cloze (n=28)	1.93 (2.23)	3.18 (3.23)	2.86 (3.28)
Editing (n=28)	2.64 (2.42)	3.43 (3.07)	3.39 (3.32)

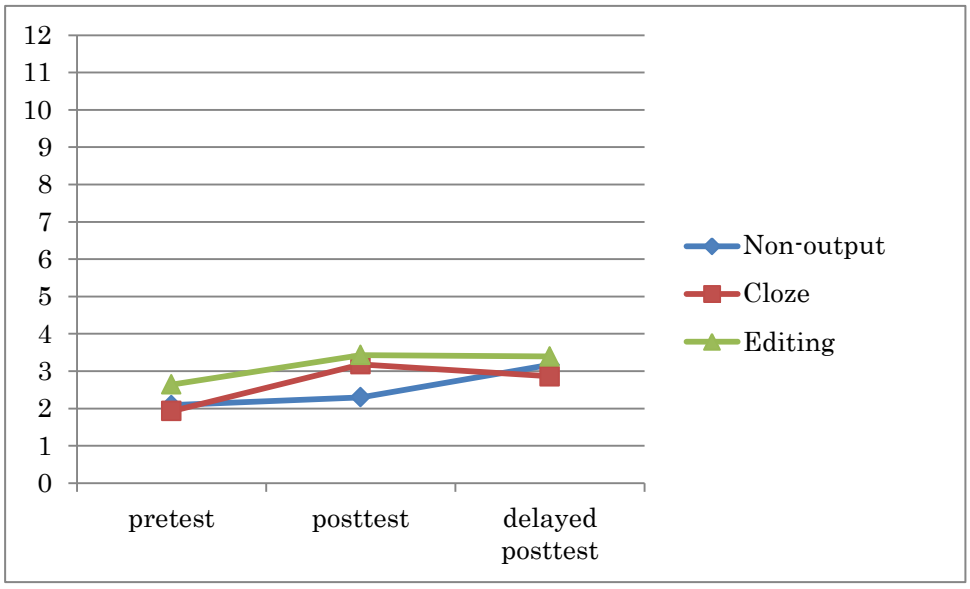


Figure 4.6 Increase in numbers of attempts in three written output

Again, one-way ANOVA results revealed that there was not any significant difference found in between group comparison, with F (2, 76) = 0.74, p = .48 for pretest comparison, F (2, 76) = 0.88, p = .42 for posttest and F (2, 76) = .20, p = .82 for delayed-posttest.

The scores were then submitted to one-way repeated-measures ANOVA and found that from pretest to posttest, both output group managed to make significant improvement (p = .002 for cloze group and p = .04 for editing group), whereas non-output group failed to do so (p = .56), which indicates that immediately after the treatment, both output groups demonstrated more effort in producing passive forms than in their pretest.

However, both output groups failed to make improvements in the delayed posttest (p = .34 for the cloze group and p = .94 for the editing group) while the non-output group managed to do so (p = .04). This is because the non-output group did not make significant improvements in posttest, so the scores remained low and that enabled the gap between these two scores to be significant. That is why it also showed significant improvement from pretest to delayed-posttest (p = .02), while only one of the output groups, the cloze group (p = .02) achieved the same. The editing group failed to show improvement because the pretest scores were already very high compare to the other two groups' scores.

On the whole, the results in this analysis were much similar to that of the written output scores. Both output groups failed to show advantage over non-output group in either their correct

production and attempts in producing the target form. However, both output groups were able to increase the effort in producing passive forms immediately after they were engaged in respective output tasks, while the non-output group failed to do so, which indicates that though output tasks did not show an advantage over non-output task statistically, they did show a facilitating effect on noticing target form and consequently led to more efforts in producing target forms immediately after the participants carried out the tasks.

As can be noted, the standard deviations have been higher than the mean scores throughout the analyses of both parts of scores and attempts. Thus, in depth investigation into individual's task performance is needed to provide more detailed explanation about the facilitating effect of output tasks.

4.4 Correlation analyses 1: Individual differences and noticing

A Pearson's *r* correlation was conducted in two different levels to examine whether there was any statistical correlation between language analytical ability and language proficiency test scores and the noticing. Since noticing is measured through two underlining scores (before and after the tasks) in present study, the relationship between individual differences and noticing is addressed by correlating language analytical ability test scores, language proficiency scores and two underlining scores.

Table 4.9 Means and SDs of language analytical ability and language proficiency test scores of three groups

	LA	LP
	Mean (SD)	Mean (SD)
Non-output (n= 23)	7.30 (2.88)	55.61 (17.74)
Cloze (n= 28)	7.43 (2.90)	56.54 (13.66)
Editing (n= 28)	7.54 (3.06)	57.25 (13.23)

Note. LA= language analytical ability; LP= language proficiency.

The two different levels here refer to the fact that the correlation analysis was done with all

participants from the three groups first, which is called all groups correlation for the convenience here, then, it was carried out with scores of each group respectively, which is called individual group correlation.

At all groups level, no statistical correlation was found between language analytical test scores and noticing, whereas statistical correlations were found between language proficiency and the two underlining scores: r (79) = .24, p < .05 with the first underlining scores; and r (79) = .23, p < .05 with the second underlining. This suggests that in general, it is language proficiency rather than language analytical ability that seems to influence learners' ability to notice the target from. To examine whether this is true at the level of individual group analysis, the Pearson's r correlation was conducted again, within individual group scores.

The results showed: there was a correlation between language analytical ability test scores and first underlining scores (r (23) = .42, p <.05) for non-output group; no correlations were observed in the cloze group; for the editing group there was a correlation between language proficiency and first underlining scores (r (28) = .44, p < .05). In sum, there was no significant correlation between individual difference and noticing at the individual group level.

4.5 Correlation analysis 2: Individual differences and acquisition

Acquisition of the passive forms has been addressed through the scores of the two types of tests (i.e., multiple choice test and written output test) which have been administered three times (i.e., pre, post and delayed-posttest) in previous analyses. Therefore, in this section, in order to investigate the relationship between individual differences and acquisition, the correlation between language analytical ability test scores, language proficiency scores and the test scores were examined at the same two different levels in the previous section.

At all groups level, language analytical ability test scores were correlated to all the test scores, except with the multiple-choice posttest scores. The results are displayed in following table. However, there were strong correlations between language proficiency and all test scores, all at significance level of 1%. This indicates the importance of both language analytical ability and language proficiency, especially that of the latter one which displayed stronger correlation. Though one may argue that there is nothing surprising about this result, yet, the relationship

between individual differences such as the two focused in present study (i.e., language proficiency and language analytical ability) and the output triggered noticing and subsequent acquisition have not been adequately researched in previous studies.

Table 4.10 Correlation between individual differences and test scores with all groups

	LA	LP	MC Pre	MC Post	MC De-post	WO Pre	WO Post	WO De-post
LA	1	.369**	.263*	0.141	.279*	.294**	.267*	.347**
LP		1	.329**	.361**	.497**	.393**	.393**	.419**
MC Pre			1	.483**	.546**	.353**	.354**	.410**
MC Post				1	.521**	.394**	.351**	.385**
MC De-post					1	.444**	.437**	.406**
WO Pre						1	.860**	.715**
WO Post							1	.843**
WO De-post								1

Note. LA= language analytical ability; LP= language proficiency; MC Pre= multiple-choice pretest; MC Post= multiple-choice posttest; MC De-post= multiple-choice delayed-posttest; WO Pre= written output pretest; WO Post= written output posttest; WO De-post= written output delayed-posttest. **: $p < .01$, *: $p < .05$

Whether the same patterns of correlations were available in each individual group was also investigated. The results varied from group to group and did not show any similar patterns. While in the non-output group, only language proficiency showed a correlation with multiple-choice posttest(r (23) = .48, $p < .05$) and delayed-posttest (r (23) = .55, $p < .01$), there were strong correlations between language proficiency and all tests scores at the significance level of 1% in the cloze group. Correlation with multiple-choice tests were: r (28) = .55, $p < .01$ for pretest; r (28) = .53, $p < .01$ for posttest; r (28) = .70, $p < .01$ for delayed-posttest; and with written output tests were r (28) = .55, $p < .01$ for pretest; r (28) = .57, $p < .01$ for posttest; r (28) = .62, $p < .01$ for delayed-posttest, which were quite similar to that of all groups correlation results.

Moreover, language ability test scores also appeared to be correlated with written output pretest scores (r (28) = .43, $p < .05$) of this group. Interestingly, for the editing group, language proficiency was not correlated with any of the test scores and only language analytical ability appeared to be correlated with multiple-choice posttest (r (28) = .44, $p < .05$) and written output delayed-posttest (r (28) = .46, $p < .05$).

To summarize, in general, both language analytical ability and language proficiency showed correlation with tests scores, especially language proficiency showed stronger correlation than language analytical ability scores, which indicates that individual differences should be considered as having influence over learners' acquisition of the target linguistic forms in a second language. In other words, this indicates that learners with higher language analytical ability and higher proficiency level tend to acquire a second language more efficiently. However, as has been reported earlier, at the individual group level, the findings were varied greatly and it is hard to make a conclusion based on those findings.

Therefore, in search of looking into the relationship between individual differences and noticing and acquisition in an individual group level, a further analysis is necessary. In the following section, each group was further divided based on their language analytical ability test scores and language proficiency scores into two subgroups to see if there is any evidence to be found.

4.6 Language analytical ability, noticing and acquisition

Since the total point for language analytical ability test is 14, participants whose scores were higher than 10 were divided into a high group and those who scored below 6 points were considered as a low group. In doing so, there are now two sub-groups in each group. A one-way between groups ANOVA was conducted with language analytical ability scores as group difference for each group. Here, the results are reported group by group. First, the comparison within non-output group revealed that the high group failed to outperform the low group in both noticing and acquisition. Second, in the cloze group, the high language analytical ability group outperformed the low group in only written output pretest (F (1, 14) = 6.05, p = .03), but not in other scores. Finally, in the editing group, the comparison found significant difference between multiple-choice posttest (F (1, 24) = 5.89, p = .02) and written output test pretest (F (1, 24) = 4.97, p = .04) and delayed posttest (F (1, 24) = 5.59, p = .03). However, no difference was found with regard to noticing.

These finding suggests that higher language analytical ability may not necessarily lead to better noticing and acquisition. Although, those with higher language analytical ability scores in editing

group did displayed better performance in two of the posttests over those with lower scores, which may indicate higher language analytical ability favored the specific task type, in this case, the editing task. Thus, in order to see whether there is a task type difference effect on noticing and acquisition combined with the language analytical ability, the comparison was made between high language ability groups in the non-output group, the cloze group and the editing group. The results yielded no beneficial effect of a specific task type on both noticing and acquisition.

4.7 Language proficiency, noticing and acquisition

Based on the participants' end of semester test scores, those who scored higher than 65 points were divided into a high proficiency group and those who scored below 50 points were put into a low proficiency group in each group respectively. The scores were submitted to the one-way ANOVA analyses in order to examine the statistical difference. The results are reported group by group. For the non-output group, the high proficiency group outperformed the low proficiency group in multiple-choice posttest (F (1, 13) = 7.74, p = .02) and three written output tests (F (1, 13) = 5.49, p = .04 for pretest; F (1, 13) = 6.35, p = .03 for posttest; and F (1, 13) = 8.61, p = .01 for delayed posttest). However, there was no significant difference with regard to noticing. For the cloze group, in terms of acquisition, the higher proficiency group outperformed the lower group at all test categories, that is, in both types of test and over all three times. However, no such difference was found in terms of noticing. The findings are summarized in following table.

Table 4.11 Summary of the one-way ANOVA analysis results for cloze group

	MC Pre	MC post	MC De-post	WO Pre	WO Post	WO De-post
sig	.006	.001	<.0001	.006	.005	.001
F	9.83	14.37	24.99	9.89	10.45	15.02
df	1	1	1	1	1	1
n	17	17	17	17	17	17

Note. MC Pre= multiple-choice pretest; MC Post= multiple-choice posttest; MC De-post= multiple-choice delayed-posttest; WO Pre= written output pretest; WO Post= written output

posttest; WO De-post= written output delayed-posttest; sig= significance level; F = F value; df = degree of freedom;

The results suggest that for the cloze group, unlike the previous findings of language analytical ability, the language proficiency obviously played a role in facilitating acquisition of the passive forms, though, the same facilitative effect did not appear with regard to noticing.

Finally, for the editing group, the significant difference was found in first underlining scores (F (1, 17) = 7.38, p = .01) and in written output retest (F (1, 17) = 6.48, p = .02). However, it should be noted that the significance level in written output posttest and delayed-posttest were both p = .06, which, though not high enough to be considered significant, was close to the 5% level and could be considered as showing some tendency toward significance. Anyway, in this group, the language proficiency did not show a strong influence as it did in the cloze group. Compared to the language analytical ability, language proficiency seemed to have minor role in influencing both noticing and acquisition in editing group. This seems to suggest that, language proficiency also favors certain particular task types, and in this case, it apparently is the cloze group.

4.8 Summary of the chapter

4.8.1 Output and noticing

In answering the first research question, both output groups outperformed the non-output group in second underling scores, which demonstrated the triggering effect of output tasks, and consequently succeeded in confirming the noticing function of output hypothesis. However, in search of the question that concerned whether there is a task type effect on noticing, although both the cloze and editing task triggered the noticing, they failed to show advantage over one another except at the first attempts at first output stage. Thus, in triggering noticing, there is partial support for the task type difference. Finally, regarding whether individual differences such as language analytical ability and language proficiency have a role in influencing noticing, the answer seemed mixed.

At all groups level, language proficiency showed a correlation with noticing while language

analytical ability did not. When broken down to the individual group level, there was no convincing evidence to be found. Then, when each individual group was further divided into high low groups according to the language analytical ability test scores and proficiency test scores respectively, again no convincing evidence was found in favor of either of the two.

4.8.2 Output and acquisition

With regard to the second research question, whether output tasks facilitate acquisition of the target form, it was addressed through examining whether there was any improvement in scores of two types of tests: multiple choice test for measuring receptive knowledge and written output test for measuring productive knowledge of target form respectively.

The tests were administered before and after the treatment: pretest, posttest and delayed posttest to show the lasting effect of output tasks. The results of multiple-choice tests showed no significant improvement over time in any group, whereas the results of written output tests were mixed.

In posttest stage which was completed immediately after the treatment section, the cloze group managed to make significant improvement compared to pretest scores (p = .003). While the non-output (p = .11) and editing group (p = .07) failed to do so. However, the cloze group failed to extend this advantage to delayed-posttest stage, where non-output group improved significantly (p = .02). The editing group failed to show improvement in this stage as well.

Comparing pretest scores with delayed-posttest scores, significant improvement were found in the non-output group (p = .004) and cloze group (p = .009), while the editing group failed again (p = .12). This indicates that, although output tasks, particularly, the cloze task did facilitate the acquisition of the target form immediately after the treatment, it failed to maintain the same effect in a longer duration. On the other hand, the non-output task, true or false reading comprehension task in this study, though failed to show immediate effect in posttest stage, managed to improve in longer run.

Although the cloze task showed more beneficial effect in posttest and delayed-posttest unlike editing task, the results of statistical analysis yielded no advantage of this task over editing tasks in both posttest and delayed-posttest scores. This finding failed to show task type effect in

promoting acquisition of the target form.

The results of correlation and further analyses of individual differences have come up with several mixed findings. Compare to language analytical ability, language proficiency showed overall stronger correlation with all the test scores.

At the individual group level, for non-output and the cloze group, especially the cloze group, language proficiency showed strong correlation with all test scores. However, it failed to correlate with any test scores in editing group. Rather, it is language analytical ability that showed correlation in the editing group with two of its posttest scores, while it failed to show any correlation in other two groups.

Further analyses of high low grouping within each group based on scores of the language analytical ability test and proficiency scores revealed similar tendency. While those with higher language proficiency in the cloze group constantly outperformed those with lower language proficiency in every tests, the same pattern was not demonstrated in non-output group and editing group. Similarly, while participants with higher language analytical ability tend to outperform those with lower language analytical ability in the editing group in multiple-choice posttest and written output delayed-posttest, similar pattern was not confirmed in the other two groups.

In general, drawing on the above finding, it can be concluded that individual differences might have a role in influencing the effect of acquisition and different individual difference favors different task type. In this study, language proficiency appeared to have more beneficial effect on acquisition when it is combined with cloze task and language analytical ability appeared to have effect on acquisition when it is combined with editing task. Meanwhile, for non-output group it is language proficiency that has a more important role than language analytical ability.

Chapter 5 Discussion

5.1 Overview of the chapter

In this chapter, the major findings of the study are discussed in light of the two main research questions along with their sub-research questions in relation to theory and research viewed in chapter 2. In this way, the chapter falls into four big parts. The first part addresses the findings concerning noticing triggered by output tasks. Second part deals with the findings concerning the facilitative effect of output tasks based on the results of two different tests. A general discussion in an attempt to address the relationship between noticing and acquisition is presented at the end, followed by the summary of the chapter.

5.2 Output and noticing

The first main research question is concerned with whether output tasks promote noticing of the target linguistic forms in subsequent input. The comparison of the first and second underlining scores was responsible for answering this question. The result yielded positive evidence to confirm the hypothesis that output tasks did actually promote greater amount of noticing than the non-output task. The first underlining scores showed that the noticing scores were equal among all the groups, output or non-output. However, after being engaged in the required activities, the underling scores on the subsequent input (the second underlining scores) showed that scores of both output groups were higher than that of non-output group. Thus unlike many previous studies (e.g., Borjigin et al. 2013; Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al, 1999; Song & Suh, 2008; Uggen, 2012), the present study demonstrated that output triggered noticing and therefore, lent support to noticing function of output hypothesis.

One of the reasons responsible for failure of output group to outperform the comparison group on noticing scores in previous studies has been that it was not because output group did not improve, it was because the comparison groups in those studies also improved on their noticing scores (see Izumi & Bigelow 2000; Izumi et al, 1999;). They argued that it was because the comprehension questions the comparison group answered have drawn the participants' attention to

target linguistic form. Since the linguistic form targeted in those studies were past hypothetical counterfactual which requires learners to pay closer attention to the forms in order to work out the meaning. Thus, doing comprehension questions was as effective to promote noticing of the form as output tasks.

In Song & Suh (2008), they found that second noticing scores of their two output groups had even slightly decreased though none of the decrease was statistically significant. They suggested that during reconstruction some linguistic features other than the target linguistic form (i.e., past hypothetical counterfactual) may have diverted participants' attention away from the targeted structure (p. 306).

The present study operationalized the two types of reconstruction in such way that, only target form and one or two distracters were missing (in cloze task) or erroneous (in editing task) in each sentence, other irrelevant features such as content words and other linguistic features were provided. This may have freed up participants' attentional capacity and thus enabled them to focus their attention more to the target forms in subsequent input and consequently led to their improved noticing scores.

Based on the findings from previous studies and the current study, it can be argued that when designing an output task, especially if there is a specific linguistic form in focus, it would be preferable to keep in mind that in order to maximize the chance of learners to direct their attention to the target form, the irrelevant linguistic elements should be kept at minimum level. This accords with what Schmidt (1990) suggested, "Task demands are a powerful determinant of what is noticed" (p.143).

Although noticing is captured through the underlining scores, one may argue that learners do not necessarily underline everything they attend to. In this case, the results of their task sheet scores were also analyzed to provide a more in depth look into the nature of the noticing triggered by output tasks. As can be known from the mean scores of both output groups in their task sheet gains, both have managed to make a statically significant improvement in their second output (cloze group 1.54 to 4.79 and editing group from 2.36 to 5.36 respectively). This indicates that the participants not only noticed more passive forms during their second input, which took place right after the first output, they also incorporated what they have noticed into their subsequent output activities. The numbers of attempts they made during first and second output (3 to 7.14 for the

cloze group and 5 to 7.61 for the editing group) also confirmed the findings that they made significantly more times of attempts to reconstruct the passive related items, such as auxiliary *be* and past participles of the verbs, though some of their attempts have not been successful.

Therefore, the findings of the present study showed the facilitative effect of output tasks on triggering noticing and at the same time provided evidence that task demands play a role in mediating the effect of output tasks on noticing. Whether there are some other factors such as task type and individual differences can influence relative efficacy of output tasks on noticing is discussed in the following two sections.

5.2.1 Task Types and noticing

One of the sub research questions aimed to find out whether a particular output task type has advantage over another. Put simply, whether it is a cloze task or an editing task that is more effective in triggering noticing of the passive form in this study. In answering this research question, comparison of the second underlining scores of these two groups was conducted, to examine whether there was a statistically significant difference between these scores. The result showed that actually the cloze reconstruction task generated slightly more underlining of the target forms than the editing reconstruction task (5.04 to 4.36), but the advantage was not statistically significant. Therefore, the two output tasks in current study appeared to be equally efficient in promoting noticing.

One possible explanation for this result can be the similarity of the two tasks. They both require participants to reconstruct same linguistic features, either in the form of filling in the missing parts (cloze task) or correction of the erroneous parts (editing task). Thus, it is natural that participants from both groups would tend to direct their attention to same parts when they were reading in the second input phase.

However, although no significant difference was found between noticing scores, closer examination of the task sheet evaluation revealed that actually there was a significant difference between numbers of attempts which took place during their first output phase. The participants from the cloze group made 3 attempts to reconstruct the target form on average, while those from the editing group made 5 attempts on average and the significance level was $p = .005$. Number of

attempts includes both successful and unsuccessful reconstruction of the target form. According to this result, participants who were engaged in editing task actually made more attempts to reconstruct the target form than those from cloze group during the first output. This indicates that the editing task directed participants' attention to target form more efficiently than the cloze task despite the fact that both tasks generated equal attention to form during the preceded reading activities. In other words, editing task was more effective in generating attention to target form than the cloze task.

However, when they carried out the tasks for the second time, there was no significant difference found on the numbers of attempts (7.14 for cloze group and 7.61 for editing group). Both group improved significantly from their first time scores, but the editing group failed to outperform the cloze group this time though scored slightly higher. This finding shows that cloze activity may be not as successful as editing tasks to draw learner attention to target form at the first try, but once learners tried out the task, they would realize what they need to know in order to successfully complete it when they have second chance.

This explains their better scores on second underlining during the second input that took place right after their first reconstruction activity (5.04 for cloze group and 4.36 for editing group). This improvement in their noticing of the target form during second input led to their improvement of the numbers of attempts during the second time of output since it followed immediately after the second input.

On the other hand, the editing group also showed the same tendency, but failed to outperform the cloze group. Since the editing group demonstrated better effort in the first time of output, one may expect them to perform better than the cloze group on both second underlining scores and second time of attempts. It should be noted that although they made more attempts to reconstruct the target form than cloze group during first output, they were not successful in converting all these attempts into correct reconstruction. The successful reconstruction of the target form was captured as the task performance score and since the two tasks were evaluated on different criteria, it was inappropriate to make direct comparison of the scores. However, since the total scores on the target form were the same 12 points, the percentage of how much they achieved out of the total score might be able to provide closer insight in the difference of the two tasks. The successful percentage of target form reconstruction of the cloze group was 13% during the first

output and that of the editing group was 20% and those of their second time performance were 40% for the cloze group and 45% for editing group respectively.

It can be observed that the editing group reconstructed the target form more successfully in both first and second time of output, and although the advantage appeared to be weak, especially the second time scores were only slightly above the cloze group, it did show a tendency to be more effective than the cloze group in terms of directing participants' attention to target form during output phases, though not in input phases. The similar findings were reported by Nassaji & Tian (2010) in their study about collaborative and individual task effect on acquisition of phrasal verb. They found that editing reconstruction task were more effective than cloze reconstruction in generating more attention to forms when it was completed collaboratively by pairs. Based on the analysis of the written transcriptions of the learner interaction during their collaborate work, it was found that working on the editing task in pairs have generated more instances of form-focused talk and feedback than the cloze task (p. 413).

In the present study, the editing task appeared to be more efficient in directing learners' attention to the target form only when they tried to reconstruct. When it came to underling during reading activity, they did not show the similar benefit, or rather tended to be outperformed by the cloze task. This may have been due to the different characteristics of the two tasks. The participants can find out the erroneous parts and somehow correct it in the editing task even without reading the input passage, while for the cloze task, without reading the input passage, it would be impossible for them to figure out the missing parts on their own. Thus, the cloze task may be more closely related to the subsequent input activity than the editing task and this observation can be supported by the higher scores of cloze group on second underlining scores than editing group, although the difference was not statistically significant.

Therefore, in order to gain a closer insight into how the cloze task actually works, it is necessary to examine it from different perspectives. When the participants carried out the cloze task for the first time, their scores were low. Therefore, they would pay more attention to the missing parts during the subsequent underlining, and it was proved so since they made significant improvement from their first underlining scores to the second one. Then, it is reasonable to assume that those who score higher during second underlining will show greater scores on their subsequent output. In other words, the first output activity pushed them to realize what they need

to know in order to carry out the task again, and those who managed to notice more of those necessary parts during subsequent the reading activity will demonstrate more attempts when they reconstruct again. Put simply, there should be a strong correlation between their second underlining scores and second attempts. The numbers of attempts include both correct and incorrect form of target linguistic form. The incorrect forms were included because they were also demonstrations of noticing of the target form. The result of Pearson's correlation analysis revealed that the correlation was significant at the level of 5% (r (28) = .43, p < .05). In short, cloze task tends to demonstrate a stronger link between second input and second output.

Editing task on the other hand, was not so dependent on the input passage, because participants can work out the solution on their own, so the input passage may have only served as a reference, the main purpose of their reading may have been just to find out whether their answer was correct or not. In this way, if they fail to find out the erroneous parts at the stage of first output, then may fail to look for the answer in subsequent input. This may explain why their second underlining scores were lower than that of the cloze group, because they would only underline what they have found problematic in their previous reconstruction activity. In other words, there should be a strong correlation between their attempts in first output performance and second underlining scores. Those who made more attempts to correct the erroneous parts during first output activity would gain higher scores on their subsequent underlining, and those who made less attempts would score less.

This assumption was proved to be true when Pearson's correlation analysis was conducted and the result revealed that the correlation between these two scores were at the significance level of 1% (r (28) = .57, p < .01). In sum, the editing task tends to demonstrate a stronger link between first output and second input. These findings are illustrated in following figures:

Cloze task

↓

Input 1→Output1→**Input 2→Output 2**

Editing task

↓

Input 1→**Output1→Input 2**→Output 2

Figure 5.1 Differential effects of cloze task and editing task

Of course, the figures do not mean that the two tasks differ totally from each other. Both output tasks triggered significantly more noticing of target forms than non-output group, and both groups went on to make successful incorporation of the noticed form in their subsequent output. The only two differences found were that: first, the editing group outperformed the cloze group in terms of their attempts during first output activity, which to some extend indicates its advantage over the cloze task in terms of noticing; second, while there was strong correlation between numbers of attempts made in first output activity and second underlining scores in editing group, there was no such correlation found in cloze group.

Therefore, it seems to be fair to say that participants of the editing group managed to notice more target forms because of their success in first output, while participants from the cloze group managed to notice more target form because of their failure in first output. How this difference may influence their relative effect on acquisition of the target form will be discussed in later sections along with the findings from relationship between noticing and individual differences.

5.2.2 Individual differences and noticing

Another sub research question concerning noticing was whether individual differences such as language proficiency and language analytical ability have an impact on effect of noticing. Results of correlation analyses conducted between language proficiency scores, language analytical ability test scores and the two underlining scores, especially the second underlining scores were

supposed to address this research question.

At all groups level, language proficiency was correlated with the two underlining scores, which suggests that participants with higher proficiency tend to underline more target forms, while at individual group level, no convincing evidence found in any group. On the other hand, language analytical ability failed to correlate with any scores in either level. Therefore, although it was hypothesized that learners with higher language proficiency and higher language analytical ability may notice more than those with lower proficiency level and lower language analytical ability, it was not supported according to the results of the current study. It seems the only notable finding from this analyses section is that language proficiency when it was correlated with noticing scores of all three groups together showed a tendency that it may influence learners' ability in noticing.

Similar findings were reported by Hanaoka (2007), in whose study he found that more proficient group noticed more linguistic features and incorporated more features into their revision writing than less proficient group, though same as the present study, the advantage was not statistically significant. He argued that this may have been because both groups noticed the problematic features in their writing and was successful in finding solutions to their self-identified problems and immediately incorporated what they noticed into their writing.

The same could be said for the present study. All participants, regardless of their language proficiency level and language analytical ability, noticed the same amount of target linguistic features at their first reading and underlining activity. It was during their second underlining activity that the two output groups outdid the non-output group, which indicated the benefit of output tasks on triggering noticing. However, within the two output groups, participants with high language proficiency level and high language analytical ability were unable to outperform those with low language proficiency and low language analytical ability on their second underlining scores, although high groups always scored higher than low groups, the advantages were not statistically different.

One of the reasons considering the results that higher language proficiency though showed a tendency to affect the noticing ability of the participants, failed to show statistically significant difference may have been due to the fact that, the proficiency level of the participants involved in present study may have been not so different from each other. It can be noted that the highest scores in high groups were 80 (cloze) and 76 (editing) and in low groups they were 50 (cloze) and

46 (editing). And the lowest scores of high groups were 65 (cloze) and 66(editing), of low groups were 30(cloze) and 27(editing). The difference between the lowest scores of high proficiency groups and the highest scores of low proficiency groups were within a difference of 15~20 points, which could have been considered as the same intermediate level together. Another reason may have been that even some participants from low groups scored as low as 27, there was only one or two such case in each group, majority of them scored above 35, thus even though the scores were very low, they could have been able to notice what was needed for completion of the task. On the whole, the proficiency level did show a more favorable effect on noticing ability of the participants.

As for why language analytical ability scores failed to show any correlation with any noticing scores at any level of analysis, it seems it could be explained by the findings of Sheen (2007). In her study, she set out to explore the relationship between language analytical ability and different type of corrective feedbacks. As a result, she found that language analytical ability showed a stronger correlation with metalinguistic corrective feedback, which indicates its positive role in learners' acquisition of explicit knowledge in particular. Thus, under implicit condition such as the one in the present study, it may not have a notable role compared to more explicit condition.

Although the individual difference such as language proficiency and language analytical ability did not show a strong influence on noticing, whether the same is true in terms of acquisition is discussed in next section according to the results of the study.

As a whole, the present study seems to indicate that when it is concerning the noticing of the forms, it is task type rather than individual difference that have a possible impact upon it. In other words, it may as well be argued that it is external factors such as task type that influence learners' noticing of forms than the internal factors such as language proficiency or the language analytical ability of the learners.

5.3 Output and acquisition

The second main research question was set to address whether output tasks facilitate acquisition of the target form, and if so whether this effect would last long. In testing this hypothesis, two kinds of tests were conducted three times (i.e., before treatment, immediately after treatment and

two weeks later) to provide evidence from both receptive and productive knowledge perspectives.

The results of the multiple-choice test yielded no supporting evidence for this research question. That means none of the three groups were successful in extending their scores from pretest to posttest and delayed posttest. Even though those who were in two output groups manifested a greater amount of noticing, they failed to show improvement from their pretest to both posttests.

This finding, however, was consistent with the previous studies (e.g., Izumi & Bigelow 2000; Izumi et al, 1999; Song & Suh 2008). It has been repeatedly demonstrated that though output may succeed in promoting noticing, it does not necessarily lead to better achievement in terms of acquisition. One of the possible explanations could be the low proficiency level of the participants in present study. Izumi (2002) argued that level of noticing is more crucial than amount of noticing. Put simply, in order for input to become intake, it is necessary to for the noticed parts to go through a deeper level of processing. According to Craik and Lockhart (1972), a depth of processing refers to relative degree of semantic and cognitive analysis and elaboration done in stimuli. The low proficiency level of the participants may have prevented them to process what they have noticed during the treatment session sufficient enough to convert the noticed input into intake.

Another reason responsible for this result can be the fact that in several previous studies it was found that output groups tended to demonstrate improvement on their written output tests but not on the comprehension tests (Izumi & Bigelow 2000; Song & Suh 2008). The similar results were observed in present study though the results of written output tests were somewhat mixed.

The written output test results showed that at posttest stage which took place immediately after the treatment, only the cloze group managed to make improvement, while the editing group and non-output group failed to do so. Therefore, in a way, it provided a partial support for the facilitating effect of an output task on acquisition of target linguistic form. Moreover, it should be noted that actually the editing group also improved on their posttest scores, and the significance level was at .06, which might as well be considered as showing a tendency of improvement.

This indicates that although noticing of the target forms does not result in better performance in terms of receptive knowledge, it may have a facilitative effect in production knowledge. Therefore, combining with the fact that editing group also showed an increasing tendency in their posttest scores, it is possible to argue in light of skill acquisition theory (Dekeyser,1997), that

learners tend to benefit from production practice in developing their production skill.

However, the two output groups failed to make a statistically significant improvement from posttest to delayed posttest, while non-output group was able to improve. This finding failed to provide evidence to the lasting effect of output tasks. The possible explanation for the failure of the two output groups to make significant improvement could be the high cognitive demand of the tasks. Unlike true or false task which requires only reading and comprehension check, cloze and editing tasks require participants to fill in the missing words into the blanket provided, or to spot the errors and edit the erroneous parts into correct forms. Izumi (2003) argued that 'cognitive demand of the task influences degree to which the learners allocate their attentional resources to form' (p. 190). The output tasks in present study have succeeded in drawing learners' attention to target forms, but this attention might have been only a shallow one, because they were too much absorbed in just completing the tasks. Besides, after they have filled in the missing parts or corrected the erroneous parts, they were left little time to process the forms further deeper and consequently forget them soon away. In other words, the output tasks might have been too demanding for the participants to engage their attention to target forms sufficient enough to convert them into intake.

The similar observation has been reported in Izumi et al, (1999), in which they argued that "the high memory load demanded by the tasks was probably too taxing for these participants and did not allow them to conduct careful analysis of the form….participants merely stored the form in their short-term memory and upon reproducing it, lost it altogether" (p. 444). This may explain why the two output groups have improved in immediate posttest but not in delayed-posttest.

The reasons why non-output group managed to make improvement in this stage may have been due to the less demanding nature of the true or false task which enabled the participants to process the content of the input passage more deeply. Although they may not have particularly noticed the target forms, but comprehension of the content could not be possible without understanding the structures. In this way, they might have processed the target forms more deeply. After all, depth of processing is more important than amount of noticing.

Another reason why they have improved significantly may have been due to lower scores they made at the posttest stage. The scores at that stage were much lower than two output groups (4.78 for non-output group, and 5.50 and 4.96 for cloze and editing group respectively). In

addition, though the non-output group improved significantly at their delayed-posttest scores, the between groups comparison showed no significant difference between the scores. In other words, two output groups also improved in their scores in delayed-posttest, and the difference of their scores and that of non-output group was not statistically significant.

In general, the possible relationship between output triggered noticing and acquisition has been partially supported in the present study and thus lent support to both noticing hypothesis and output hypothesis. However, it seems output facilitates learners' acquisition of the target forms only in terms of their production skills and this facilitation failed to demonstrate an enduring effect.

5.3.1 Task type and acquisition

One of the sub research questions attempts to address if there is any differential effect of task type on acquisition of target form. Put simply, which group, cloze or editing can outperform the other on three times of both types of tests, especially those of posttest and delayed-posttest scores.

As it has been mentioned in the previous section, multiple choice test scores were not able to provide any evidence in favor of either group. Both groups failed to extend their scores in posttest and delayed-posttest, and there was no statistical difference between groups comparison. In the pretest stage, however, the editing group scored higher than the cloze group (5.21 to 4.04) and the difference was statistically significant. In addition, the mean scores of editing group in posttest and delayed-posttest have been higher than the cloze group: 5.43 to 4.25 at posttest and 5.14 to 4.57 at delayed-posttest, but the differences were not significant. In short, there was no effect of task difference in terms of acquisition of passive form based on the performance on receptive knowledge tests.

The performance on production tests on the other hand, also failed to show any different effect of task types. Despite the fact that cloze group did make a significant improvement from their pretest to posttest scores, the significance was only found within group but not between groups. In other words, although the cloze group made an improvement on their posttest scores, the improvement was not sufficient enough to outperform editing group. The posttest mean score of cloze group was 5.50 and that of the editing group was 4.96, and the difference between these two

scores appeared to be not statistically significant.

Based on the statistical analyses, the findings suggests that there was no difference between the cloze task and editing task with regard to acquisition of the target form, on both types of tests. Similar findings have been repeatedly demonstrated in previous studies on task types and relative efficacy on noticing and acquisition. In Song & Suh (2008) who explored the different effect of reconstruction task and picture-cued writing task, they found that although picture-cued writing task generated more noticing of the target forms than reconstruction task, they were not more effective than the reconstruction task when it concerned with the acquisition of the target form. Ghari & Moinzadehb (2011), who also investigated the differential effect of reconstruction task and picture-cued writing task, found that output tasks not only promoted noticing of the target forms but also promoted the acquisition of the target forms. Yet again, the two tasks did not differ in their effects on learning of the target forms. The picture-cued writing task in their study generated more noticing than reconstruction task as it did in Song & Suh (2008)'study.

Therefore, along with the finding from previous studies it seems to be fair to argue that output did promote noticing, and some output tasks are more effective in generating noticing than others (e.g., picture-cued writing task). However, in terms of acquisition, it seems output tasks tend to have a similar effect. This finding inevitably leads to the discussion of the relationship between noticing and acquisition. Empirical studies demonstrated that noticing is necessary for learning, because the output tasks generated more noticing and manifested better performance at least on production tests, this confirmed that in order for learning to take place, noticing is necessary. This will lead to argument that if noticing is necessary for learning, then more noticing should enable better learning. This hypothesis is supported partially, because learners from output groups did notice more and performed better on their production tests. Again, at this stage, one may argue that, if this is so, then the learners who carried out picture-cued writing task and noticed more forms than those who completed reconstruction task should do better on their tests, at least on their production tests. The problem lies here, when they were supposed to demonstrate better performance, they simply did not. One of the possibilities is that although one output task may promote more noticing than the other output task, this does not mean that the other one did not promote noticing. Take Song & Suh (2008) as an example, picture-cued writing task did generate more noticing than reconstruction task, but actually the reconstruction task also improved in their

second underlining scores. Therefore, even though the reconstruction task did not show greater gains on noticing, the gains might have been sufficient for them to acquire the target forms as well as the picture-cued writing task did.

Another possible explanation could be that when it comes to acquisition, noticing alone may not be able to provide a full account of its nature. As Song & Suh (2008) put it, while it is obvious noticing is necessary for learning, there might be some other more complicated cognitive process to link attention and learning (p. 307). This argument resonates with that of Izumi (2002), as he also argued that depth of noticing is more important than amount of noticing. Therefore, it is necessary to further explore what factors really matter to the link between quantity (amount) of noticing and quality (depth) of noticing.

According to this study, task type seems to obtain no such influence over this link. Tasks may differ from each other in generating different quantity of noticing, but so far, they failed to show differential effect on depth of noticing. However, one study did succeed in showing task type difference. In Nassaji & Tian (2010), they found that both individual and collaborative editing tasks led to significantly higher scores than individual and collaborative cloze tasks in terms of the learning of phrasal verbs. In analyzing transcription of participants' interaction, they found doing editing task generate more instances of form-focused talk and feedback than collaborative the cloze task. They continued to argue that "such negotiation in the editing task could have oriented the learners' attention to the phrasal verbs more effectively and hence could have resulted in deeper understanding and knowledge of the target forms"(p. 413).

This finding may suggest that editing tasks are more effective than cloze tasks because they are more form-focused, which means deeper understanding may come from more elaborate negotiation of form. In this way, a collaborative editing task could be considered as an effective pedagogical task aiming to promote learning of a certain linguistic feature. That said, the collaborative editing task may have outperformed cloze tasks, but it did not outperform the individual editing task in their study, hence suggested that pair works do not always lead to a better acquisition than the individual work. Thus, further research is needed to explore more effective way of implementing cloze and editing tasks.

In addition, a closer look at the participants who have shown higher scores on production test scores in present study showed that they are either with high language proficiency or high

language analytical ability or combination of the both. In fact, majority of them were of combination of both. Discussion on individual difference and its relative effect on acquisition will thus make up the next section.

5.3.2 Individual difference and acquisition

Another sub research question was concerning about whether language proficiency and language analytical ability may have a role in mediating the effect of output on acquisition. The correlation between language proficiency scores, language analytical ability test scores and all the test scores (i.e., pretest, posttest and delayed-posttest scores of both types of tests) along with the comparison analysis within each group were supposed to address this question.

The correlation results revealed that at all group level, both language proficiency and language analytical ability are correlated with all the tests scores. This finding conforms to the anticipation that learners with higher language proficiency and language analytical ability tend to perform better on their test scores, which in turn suggests that they achieve better acquisition of the target forms.

However, the main interest of the research question lies in whether there is any role of individual differences on link between output and acquisition. Thus, the analysis was conducted at individual group level in attempt to address this issue. The results were reported here group by group first and then were discussed in general. In the non-output group, correlation was found only between language proficiency scores and posttest and delayed-posttest scores of multiple-choice test. In the cloze group, correlation were found between language proficiency scores and all test scores, which means, pretest, posttest and delayed-posttest scores of both multiple-choice test and written output test. Language analytical ability scores were correlated only with pretest scores of written output test. In the editing group on the other hand, there was correlation between language analytical ability test scores and posttest scores of multiple-choice test and delayed-posttest scores of written output test. Language proficiency was not correlated with any of the scores of either test.

Therefore, it turned out that while language analytical ability showed correlation with the editing group in posttest scores (multiple choice test posttest scores and written output test

delayed-posttest scores), and the cloze group (pretest scores of written output test), language proficiency showed correlation with non-output group (two posttest scores of multiple-choice test) and more strongly with cloze group (all of the scores of both types of tests).

Based on these results, it can be observed that for the non-output group, language proficiency seems to have a beneficial influence on acquisition of receptive knowledge of the target forms, while language analytical ability seems to have no such influence on either receptive or productive knowledge at all. In the two output groups, the patterns seems rather clear that while language proficiency tends to be more influential on effect of acquisition of both receptive and productive knowledge in the cloze group, it appears to be of no such importance in editing group, due to the fact that it correlated with none of the scores of the editing group. However, language analytical ability on the other hand, seems to be influential on the effect of acquisition of receptive and production knowledge of participants in the editing group, though it failed to show such effect in the cloze group.

These findings are confirmed again by the comparison analyses conducted within each group. In order to test whether the results of correlation analyses were reliable, each group was further divided into two sub groups based on their language proficiency scores and language analytical ability scores on separate occasions. That means, first each group was divided into two sub groups based on their language proficiency to see whether more proficient participants really outperform less proficient ones on their test gains. The results showed that while more proficient participants outperformed less proficient ones on each test scores, the similar effect was not found in non-output group and editing group. Then, when each group was divided into two sub groups based on their language analytical ability scores, participants with high language analytical ability outperformed those with low analytical ability on the multiple-choice posttest and written output delayed-posttest in editing group, but no results of same effect were observed in other two groups.

On the whole, it seems fair to argue based on these results that, different individual differences seem to differ in their influence over different task types and their relative effect on acquisition. To put it simply, learners with higher language proficiency tends to be benefitted more from the cloze task than those with lower proficiency level and learners with higher language analytical ability tend to benefit more from editing task than those with lower language analytical ability.

In fact, similar findings have been reported in previous studies. Sheen (2007) found that language analytical ability was more strongly related to error correction test results because error correction requires more metalinguistic knowledge. On the other hand, the cloze task and reading comprehension task such as true or false appear to require language proficiency rather than analytical ability.

This may suggest that if teachers plan to implement a cloze task in their classroom, they should consider the proficiency level of their students. If their students' proficiency level is not so high, they may not be able to benefit from doing the task. The teachers may need to work out other solutions to adjust the task, such as reducing the task demand or perhaps asking the students to carry out the task in pairs or in groups. Similarly, if they try to implement an editing task, perhaps it is necessary for them to keep in mind the language analytical ability of their students.

5.4 Relationship between noticing and acquisition

Schmidt (1990) has argued that noticing is the necessary and sufficient condition for converting input to intake. However, he later modified this claim to learning without noticing may be possible, but noticing is beneficial (Schmidt, 1994). So it is worthwhile to look at the link between output triggered noticing and acquisition, to see whether noticing triggered by output activities lead to better acquisition of the form (second main research question in present study) and if there is no favorable evidence for it, then, whether acquisition without such noticing is possible.

This study seems to have provided some partial support about the beneficial effect of noticing on second language acquisition. Namely, the two output groups who demonstrated better noticing than the non-output group also managed to improve, or at least showed a tendency of improvement (i.e., editing group), in their performance from pretest to posttest in production test which took place immediately after the treatment. Though the improvement did not necessarily outperformed the non-output group, still, it indicated the facilitating effect of noticing on the acquisition to some extent and lent support to Schmidt' noticing hypothesis (1990, 2001).

In discussion of the link between noticing and acquisition, it has been demonstrated in Izumi (2002) that depth of noticing was far more important than the amount of noticing in order for

input to be converted to intake and acquisition. In his study, Izumi found that despite the fact that subjects engaged in input enhancement treatment they noticed the target form (English relative clause) more than the subjects carried out the output task (reconstruction task), they were outperformed by those from output group on both comprehension and production tests. Besides, output groups in his study they also outperformed input group who processed the input for sole purpose of comprehension. These findings led him to arrive at the conclusion that compared to external attention drawing device such as input enhancement, output being an internal priming device promoted the learning better. He continued to point out that by producing output, learners are provided with chance to compare their interlanguage and target language, and in doing so have had the opportunity to replace imperfect forms in their interlanguage with correct forms in target language. In contrast, the input enhancement, though successful in drawing learner attention to forms, may not have necessarily pushed them to further process the noticed forms (p. 567).

The findings of the present study, on the other hand, though lend partial support for noticing function of output hypothesis and noticing hypothesis; do not fully echo with what Izumi (2002) has found. The two output groups failed to improve on their receptive tests, and although were able to improve on their production test scores, the improvement were not good enough to outperform the non-output group. However, the correlation analyses results in present study seem to indicate that individual differences may have a mediating role on link between noticing and acquisition.

Since the statistical analyses could provide no further accounts on this issue, further small scale analyses were conducted in the hope of addressing the relationship between noticing and acquisition. Due to the fact that no groups, output or non-output, have improved in their receptive test scores, production test scores were thus considered more preferable to be taken into account for further analyses.

To address the relationship between noticing and acquisition, it is inevitable to revisit the integrated model of SLA proposed by Gass (1988). Therefore, the discussion below is going to be organized along the stages Gass proposed in her model.

In Gass's integrated model of SLA, there were two stages before the stage of intake: apperceived input and comprehended input. It can be assumed that when the participants of the present study were exposed to the input for the first time, they were going through the

apperception stage in this model. According to Gass (1988), at this stage what the participants notice from in the input would be influenced by factors such as the frequency of certain linguistic features and the existing knowledge of the target language. As it turned out in present study, at this stage the underlining scores of the participants correlated with their language proficiency scores. This may indicate that existing knowledge about the target language does have an impact on what linguistic features the learners pick up from the ambient input. The underlining scores, at this stage, were equal across the groups.

However, this stage does not yet lead to intake stage. Learners have to have an opportunity to compare what they noticed (target forms) and what they already know (existing knowledge in their IL), and in doing so, they undergo what Gass (1988) calls it, the comprehended input stage. In the present study, it could be considered that this stage took place when the participants were exposed to the input for the second time after they were engaged in different tasks. The intervening task between two input passages is expected to enable learners to notice the gap between what they have just noticed and what they need to know to complete the task successfully. Therefore, after carrying out the task for the first time, they are expected to notice this gap and consequently seek for the required information in the following input. As it is hypothesized in output hypothesis, the output tasks will enable learners to notice the gap and accordingly pay more attention to linguistic features required in the subsequent input. This is exactly what happened to the participants of the two output groups in present study. After doing the reconstruction tasks, they realized what they need to know to carry out the task, so they underlined the target forms more than the participants from the non-output group whose task only required them to understand the passage on the whole to carry out the comprehension check practice followed.

Gass (1988), in defining the comprehended input stage, also argued that there are different levels of comprehension of the input. She continued to argue that not all comprehended input could lead to intake. She suggested “an analysis at the level of meaning is not as useful for intake as an analysis made at the level of syntax” (p. 206). As has been mentioned earlier, Izumi (2002) argued the importance of the level of processing at this stage and his study demonstrated output tasks did enabled learners to process the target forms more sufficiently than the input enhancement group and input group. He arrived at this conclusion when he found the output

groups in his study outperformed the other groups in posttests. In other words, to figure out what level of processing learners have made at comprehended input stage, one way to capture is to examine their test performance after the treatment sessions.

A question emerges here, since mean scores of groups in present study did not yield much about whether the participants processed the target forms sufficient enough or not, the analyses hereafter will focus on those who succeeded in making improvement in their performance in posttests. The hypothesis here works in this way: if one has achieved improved performance after treatment, then he or she must have noticed more and incorporated what he or she noticed into the subsequent task performance. Moreover, since the performance of all groups on multiple-choice tests were almost the same across the different test stages, it is considered not so helpful to provide any insight into the different effect of different tasks. Thus, production test performance was analyzed instead.

First, the gain scores of the posttest were calculated in each group to see how many of them were successful in extending their scores in written output tests. The following tables show the result of such analysis of each group, and the tables are presented in the non-output group, the cloze group and the editing group orders.

Table 5.1 Written output posttest gains of non-output group

Participants	Pretest	Posttest	Gains
1	5	17	12
2	0	0	0
3	8	18	10
4	14	16	2
5	0	3	3
6	0	0	0
7	2	0	-2
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	14	24	10
14	3	0	-3
15	3	3	0
16	12	8	-4
17	0	0	0
18	0	0	0
19	0	0	0
20	14	19	5
21	2	2	0
22	0	0	0
23	0	0	0

Note: Gains=Posttest scores-Pretest scores

As can be noted from the table, there were six participants who have made improvement in their

posttest performance in this group.

Table 5.2 Production posttest gains of cloze group

Participants	Pretest	Posttest	Gains
1	0	0	0
2	4	7	3
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	2	2
8	0	4	4
9	0	0	0
10	2	0	-2
11	3	6	3
12	8	8	0
13	0	0	0
14	0	0	0
15	20	26	6
16	16	23	7
17	0	0	0
18	3	15	12
19	0	4	4
20	6	20	14
21	4	8	4
22	2	0	-2
23	3	6	3
24	12	11	-1

25	0	0	0
26	0	9	9
27	0	0	0
28	2	5	3

Note: Gains=Posttest scores-Pretest scores

In this group, 13 participants have managed to gain increase in their posttest scores compare to their pretest performance.

Table 5.3 Production posttest gains of editing group

Participants	Pretest	Posttest	Gains
1	0	3	3
2	4	0	-4
3	0	3	3
4	0	0	0
5	9	9	0
6	3	3	0
7	0	0	0
8	0	0	0
9	12	6	-6
10	17	19	2
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	2	2	0

18	11	21	10
19	2	2	0
20	11	20	9
21	2	3	1
22	15	21	6
23	0	3	3
24	0	0	0
25	3	3	0
26	0	0	0
27	5	21	16
28	0	0	0

Note: Gains=Posttest scores-Pretest scores

As can be observed, there were nine participants from this group who have been successful in extending their scores in posttest performance.

After deleting those who failed to make improvement in their posttest scores from each group, there are now six subjects left in non-output group, 13 for cloze group and nine for editing group for the further analysis.

The analyses were conducted to examine what these successful learners might provide to address the relationship between noticing and acquisition. This makes it necessary to examine the mean scores of each group in each category which includes noticing scores, test scores, and the task performance scores. The following tables show the results of these analyses.

Table 5.4 Mean noticing scores of each group

	Noticing 1	Noticing 2
	Mean (SD)	Mean (SD)
Non-output (n=6)	2.83 (2.04)	2.17 (1.47)
Cloze (n=13)	3.53 (2.50)	6.15 (3.95)
Editing(n=9)	3.89 (1.45)	4.56 (1.59)

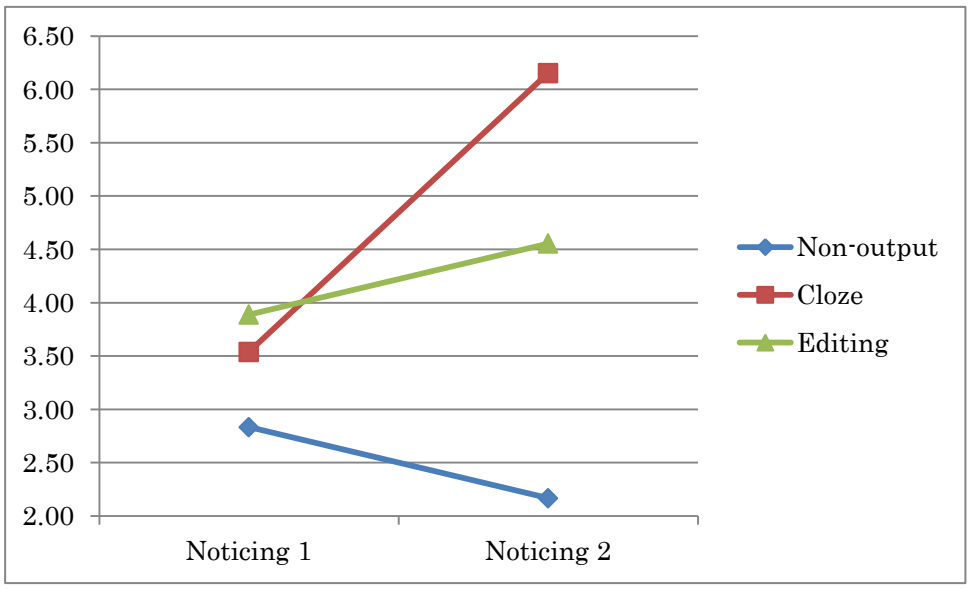


Figure 5.2 Noticing scores of each group

Only the cloze group improved significantly in their second noticing scores, while the editing group only improved slightly and non-output group did not improve at all. This finding is a bit different from what was found from the overall analyses. The overall data yielded positive evidence for both output groups. According to this analysis, the results for cloze group and non-output group remained the same whereas editing group provided a different picture from the overall data results. It seems those who made improvement in their production test in editing group did not necessarily notice more forms when they read through the input passage for the second time. This finding however, did not contradict the results of overall data analyses in which it was found that editing tasks trigger noticing in different way than the way cloze task did. In the cloze task, the gap is already provided there, what the learners need to do is to look for what

information is missing in the subsequent input activity which supposedly comes up with more noticing of the forms. The editing task on the other hand, requires learners to spot the erroneous parts on their own, in this way, only those who are able to specify the problematic features will seek the correct forms form the subsequent input. That said, again, when the cloze group demonstrated the possible relationship between noticing and acquisition, the editing task fails to achieve the same. This result also accords with that of the overall analyses. Overall data indicated that while participants from the cloze group improved on their noticing scores after treatment, they also improved on their production posttest performance which took place right after the treatment. The editing group, although demonstrated a tendency of improvement, failed to do the same.

As for the editing group, the overall data revealed that they improved in their noticing scores but not in their acquisition scores, this small scale analyses indicate the vise versa, that is, those who did improve in their acquisition did not necessarily demonstrated better noticing. However, apart from the underlining scores which are utilized as the noticing scores in analyses, there is another way of investigating noticing, that is, the attempt the participants made during carrying out the task. In the present study, the attempts are counted as number of times when participants make both successful and unsuccessful production of the target linguistic features required. Since there are 12 target related items required to be produced in each output task, the maximum number of attempts could be 12. The criteria were explained in chapter 3, but for the convenience of understanding, an actual scoring sample of a participants form the editing group is presented below:

A very exciting year

In January, Amy employed as the secretary <u>for</u> the Inter-Tel company.

by (1 point, 1 attempt)

In March, she sended <u>to</u> school with the company to study statistics and information technology.

for (0 point, 0 attempt)

In April, she was <u>giving</u> a raise.

gave (0 pint, 1 attempt)

Just two months late, she <u>promoted</u> to the position of supervisor of her department.

promisted (0 point 1 attempt)

In August, she chooses as the “Employee of the Month”.

In October, she gave another raise.

In November, she was inviting to apply for a position in the company’s overseas office on

Delete (0 point, 1 attempt)

Bangkok, Thailand.

In December, she provided the new position of Thailand.

In(1 point)

At end of month, she was award the “Employee of the Year”.

(the) (the) (2 points)

She can’t believe all the wonderful thing that has happened to her since she was hire just twelve months ago.

This student’s total scores on task performance was 4 points, 1 point out of the 4 was on target related form, and she made 4 times of attempts to correct the target related forms. As it can be observed form the task sheet that, she made 4 times of attempts to correct the erroneous parts, although only succeeded once. This failure of producing correct forms does not necessarily mean that she did not notice them. Thus, number of attempts should be regarded as complementary way of measuring noticing to underling.

The mean scores of attempts made by the editing group were in fact remarkably high each time they carry out the task compared to the cloze task as can be seen from the table below.

Table 5.5 Mean numbers of attempts and SD of two output groups

	First Attempts	Second Attempts
	Mean (SD)	Mean (SD)
Cloze (n= 13)	5.15 (2.97)	7.92 (3.45)
Editing (n= 9)	6.56 (2.35)	8.67 (2.18)

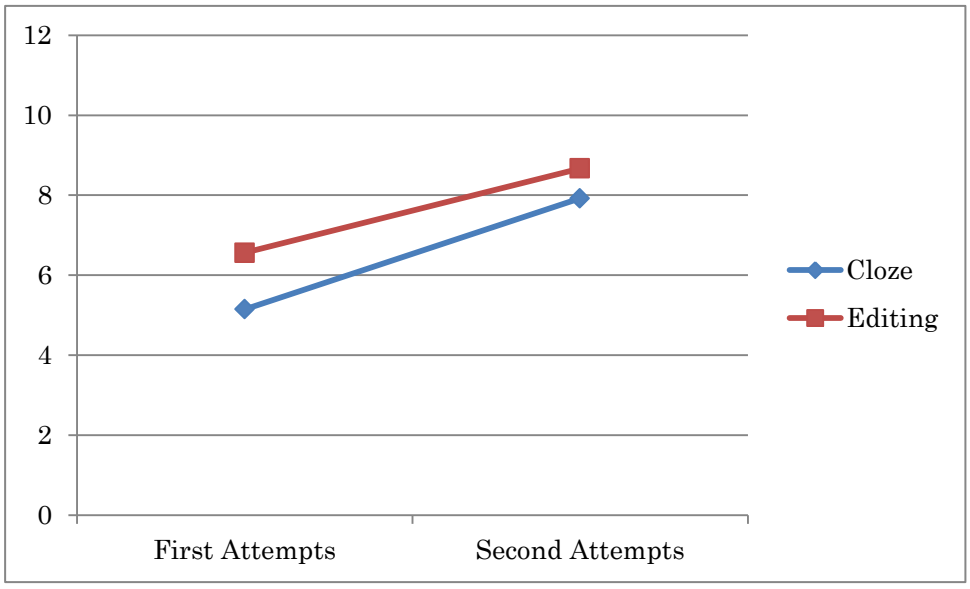


Figure 5.3 Mean numbers of attempts made by two output group during tasks

And actually editing group outperformed cloze group significantly (p = 0.057) according to the results of One-way ANOVA analyses.

Then, a question may emerge here, why they did not underline that many target related items if they have noticed the problematic features. The possible reason is because of the character of the editing task, since part of the target structure, either auxiliary *be* or a past participle is already available on the task sheet, they might just look for the correct form of either part and thus come up with half score not a full score. The criteria of scoring the underling parts have been explained in chapter 3, but for the convenience of understanding, an example is provided below. Suppose a participant has noticed the erroneous part of the following sentence:

Amy was sended to school by the company.

He or she is expected to look for the correct form in the subsequent input, in which the sentence is provided as:

Amy was sent to school by the company.

Then, instead of underling both *was* and *sent*, he or she may just end up underlining *sent* alone.

And in this situation, the noticing score he or she can get is 0.5 instead of 1 when both items were underlined. And this assumption is proved to be true if you look at the following sample. This is an actual sample of first and second underlining of a participant from editing group who made significant improvement from first production test scores to second production scores (form 5 points in the pretest and 21 in the posttest).

First time underling score: 3 points

A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company. (0.5 point)

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise. (0.5 point)

Just two months later, she was promoted to the position of supervisor of her department. (0.5 point)

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise. (0.5 point)

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago. (1 point)

Second time underling score: 3 points

A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company. (0.5 point)

In March, she was sent to school by the company to study statistics and information technology. (0.5 point)

In April, she was given a raise. (0.5 point)

Just two months later, she was promoted to the position of supervisor of her department. (0.5 point)

In August, she was chosen as the “Employee of the Month”. (0.5 point)

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand. (0.5 point)

In December, she was provided the new position in Thailand.

At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

As it shows, he did not make improvement in his noticing scores, but in fact, in the second time underlining, he noticed more target related items than first time. It is because he only underlined the past participle but not together with auxiliary *be* every time, which is considered as only constituting the half of the whole structure, and a half point is given accordingly.

Therefore, considering both the underlining scores and the attempts the participants made during the task completion process, and combining with the findings from overall data analyses, it might be fair to say that those who manifest better noticing scores may not necessarily demonstrate better performance in terms of acquisition, but those who demonstrate better acquisition of the target form, may also manifest better noticing of the forms.

In other words, the relationship between noticing and acquisition might be put in this way: noticing may not be the sufficient condition for learning to take place, but it is necessary for it to happen.

Then, as the overall data analyses did, in this small scale study it is also necessary to examine the factors such as task types, and individual differences and their possible influence over the link between noticing and acquisition.

5.4.1 Task type, noticing and acquisition

In this section the analyses are conducted to see whether among those who were able to make

improvement on their posttest scores could possibly exist a task type effect. Put simply, the purpose of the analyses in this section is to search for an answer to the following question: which task may better facilitate learning compared to the other two?

The following table summarizes the mean scores and the standard deviations of three groups' performance on the three times of written output tests and is followed by the figure that illustrates the scores.

Table 5.6 Written output test results of each group

	pretest	posttest	delayed-posttest
	Mean (SD)	Mean (SD)	Mean (SD)
Non-output (n=6)	9.17 (5.88)	16.17 (7.03)	15.33 (9.33)
Cloze (n=13)	4.69 (6.25)	10.39 (7.93)	8.85 (9.49)
Editing(n=9)	6.78 (6.82)	12.67 (9.19)	13.44 (10.36)

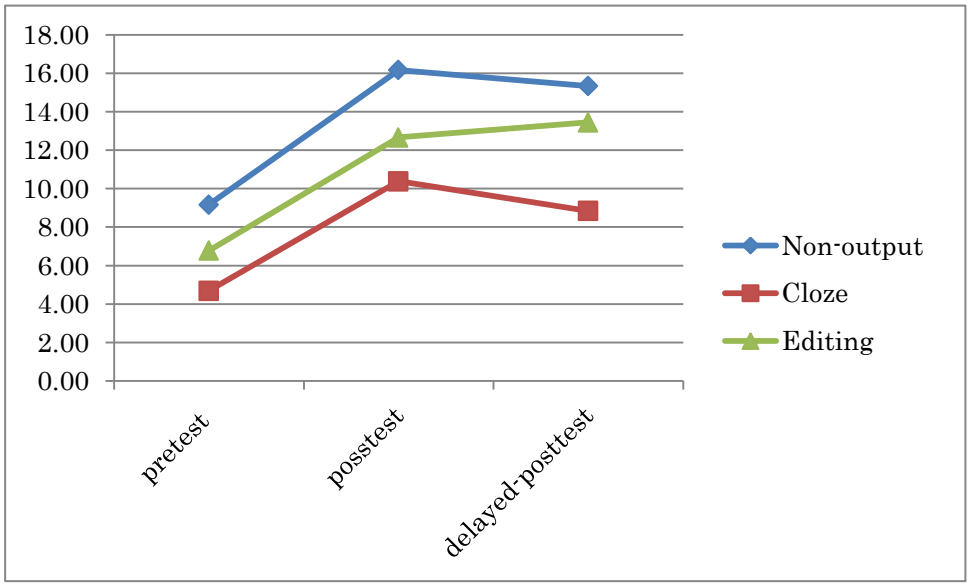


Figure 5.4 Written output test results of each group

From the figure, it is difficult to notice any relationship between noticing and acquisition, which has been explored in the previous section, since the cloze group who scored the highest in second noticing scores was not the one which scored the highest among the three on the written output

test performances. Yet, it should be noted that there were only six out of 23 participants in non-output group have made this improvement and besides, it can be observed from table 5.1 that three of these six happened to improve by over 10 points in their posttest. While in the cloze group 13 participants have come up with improvement but with varied arrange of points: some of them improved greatly while some only slightly. The similar patterns can also be observed in the editing group.

In this way, it is not fair to make any argument if it is just based on the comparison of the mean scores. Besides, there is no statistical difference between those scores. This suggests that only examining the mean scores may not amplify the judgment. A possible way to explore group difference then should rely on the qualitative analyses of these scores. As has been mentioned, 13 out of total 28 participants were able to improve on their posttest scores, and the same was achieved by nine out of total of 28 participants from the editing group and six out of 23 from non-output group. This makes the percentage of successful learners in each group 46%, 32% and 26% respectively. Learners who were engaged in the cloze task were most likely to improve, followed by the editing task and the true or false task. The gap is quite remarkable for the cloze group when it is compared to the other two. This is confirmative for the result of the overall data analyses in which only the cloze group made significant improvement on their posttest performance. The editing group also produced a slightly higher percentage of successful learners than the non-output group.

These findings seem to suggest that output tasks not only trigger better noticing of the target forms but also have a more facilitative effect on the acquisition of those forms. However, whether they carry on this effect for a longer time turns out to be less positive. The table and figure above show that these successful learners did not appear to be so successful in their delayed-posttest scores. In fact, only one out of six learners from non-output group continued to make improvement in delayed-posttest stage and two out of 13 learners from cloze group and five out of nine learners from editing group managed to do the same. The following table shows the gain.

Table 5.7 Written output delayed-posttest gains of non-output group

Participants	Posttest	Delayed-posttest	Gain
1	17	24	7
3	18	18	0
4	16	9	-7
5	3	0	-3
13	24	24	0
20	19	17	-2

Note: Gains=Delayed-posttest scores-Posttest scores

Table 5.8 Written output delayed-posttest gains of cloze group

Participants	Posttest	Delayed-posttest	Gain
2	7	0	-7
7	2	2	0
8	4	3	-1
11	6	0	-6
15	26	26	0
16	23	24	1
18	15	15	0
19	4	2	-2
20	20	20	0
21	8	12	4
23	6	6	0
26	9	0	-9
28	5	5	0

Note: Gains=Delayed-posttest scores-Posttest scores

Table 5.9 Written output delayed- posttest gains of editing group

Participants	Posttest	Delayed-posttest	Gain
1	3	0	-3
3	3	15	12
10	19	8	-11
18	21	30	9
20	20	14	-6
21	3	2	-1
22	21	24	3
23	3	6	3
27	21	22	1

Note: Gains=Delayed-posttest scores-Posttest scores

It turns out that while only one or two learners from non-output group and cloze group who have made better performance on their production posttests continued to improve in delayed-posttest stage, five learners from the editing group were able to make continuous improvement. This means more than half of those who made improvement in their posttest scores have again raised their scores in delayed-posttest stage, which took place two weeks after the treatment session and immediate posttest session.

This finding seems to suggest that in terms of long term effect, editing task may come on top of cloze task and non-output task. Due to its qualitative nature and small numbers of the subjects, it may be less convincing to claim that the editing task is more beneficial for the acquisition of the target forms in a long term, but at the same time, in a way it might indicate that the editing task might have helped learners to process the forms more deeply so as to be able to acquire them. The cloze task on the other hand, despite showing better effect on immediate learning, the effect fades away in a longer term. The non-output group not only failed to show facilitative influence on acquisition on immediate learning (compared to the two output groups), but also in a long term.

Generally speaking, two output groups are more likely to notice more target forms and accordingly more likely to achieve a better acquisition of the target forms. Within two output

groups, the cloze group seems to be more likely to improve in the immediate posttest, while the editing group seems to show a tendency of continuous improvement in a longer term.

Among those participants who continued to make improvement in delayed-posttest stage however, there seems to be a similar pattern. This pattern will naturally lead to the next discussion, because all of those who made continuous improvement appeared to be with high language analytical ability and at the same time, high proficiency level.

5.4.2 Individual differences, noticing and acquisition

In previous sections, the discussion of the results of overall data analyses suggested that in terms of noticing, only language proficiency showed a correlation with noticing scores at all groups level, but not in individual group level. Language analytical ability scores failed to show any correlation with the noticing scores at either at all group level or individual group level. While in terms of acquisition, language proficiency scores tended to correlate with test performances of cloze group, and language analytical ability scores were more inclined to correlate with posttest scores of editing group. The findings were again confirmed by the further analyses conducted between high and low sub groups within cloze and editing group respectively.

However, when considering the relationship between the noticing and acquisition, these findings seem inadequate to come up with a conclusion. To address whether individual differences such as language analytical ability and language proficiency influence the link between noticing and acquisition, it seems more reasonable to examine the characters of participants who demonstrated better acquisition of the target forms.

The following table summarizes the means and SDs of the language analytical ability scores and language proficiency scores of participants who achieved improvement on their production posttest scores in each group.

Table 5.10 Mean and SD of language analytical ability scores and language proficiency scores of each group

	LA	LP
	Mean (SD)	Mean (SD)
Non-output (n = 6)	8.00 (2.28)	63.33 (6.56)
Cloze (n = 13)	7.85 (2.79)	61.23 (11.54)
Editing (n = 9)	9.11 (3.06)	62.33 (8.03)

Note: LA= language analytical ability scores; LP= language proficiency scores

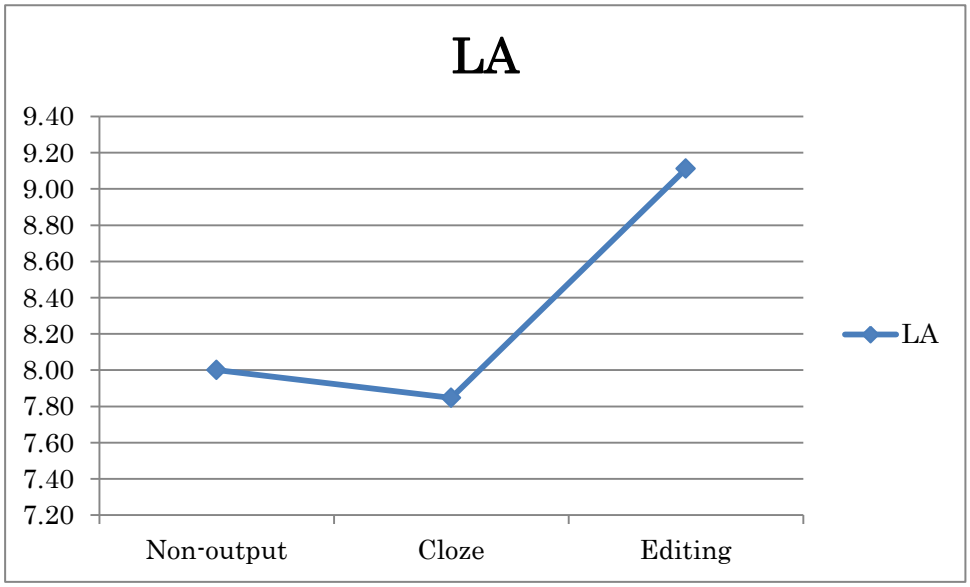


Figure 5.5 Means of language analytical ability scores of three groups

Note: LA= language analytical ability scores

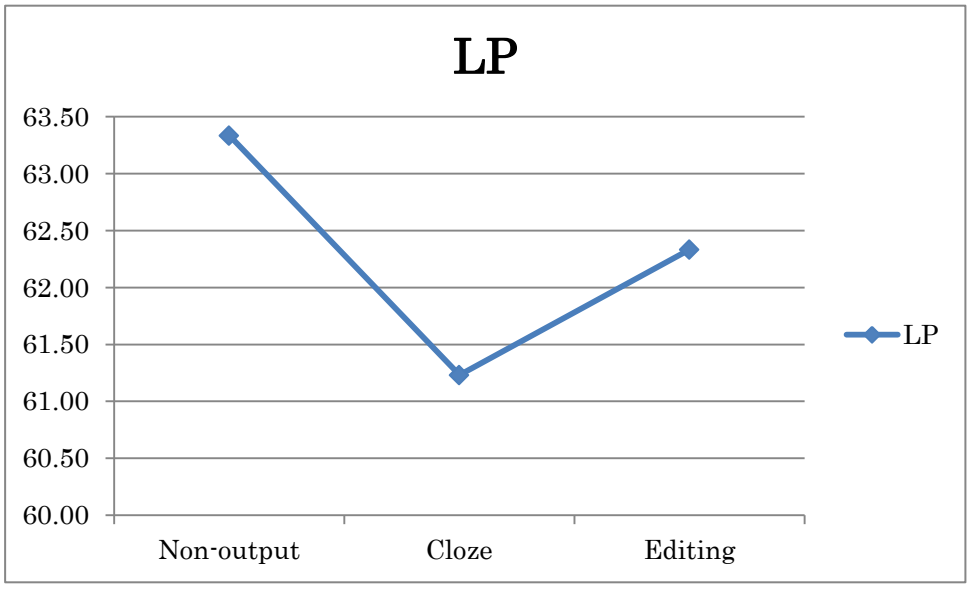


Figure 5.6 Means of language proficiency scores of three groups

Note: LP= language proficiency scores

Note that the total scores are 14 for language analytical ability test and 100 for the language proficiency test. As can be observed form the table that those who demonstrated better acquisition were those who with higher language analytical ability and language proficiency. Furthermore, the mean of language analytical ability test scores for the editing group was the highest of the three, which again confirmed the finding from overall data analysis that learners with higher language analytical ability are more likely to benefit more from the editing task.

The proficiency level of the each group appears to be equally high, which seems to suggest that in order for learners to benefit most from the tasks and learn the target form more effectively, they should possess relatively high proficiency level. However, this finding does not support the finding of the overall data analyses that proficient learners are more likely to benefit from cloze task.

With regard to the acquisition on longer term, it turns out that five out of nine learners from the editing group, only one learner form non-output group and two from cloze group continue to make improvement on their delayed-posttest scores. In order to give a full account of these successful learners, their language analytical ability scores, language proficiency scores and all the test scores are provided below.

Table 5.11 Language proficiency and language analytical ability and two noticing scores of successful learners in long term

Participant	Group	LA	LP	Noticing 1	Noticing 2
1	Non-output	10	66	3	4.5
2	Cloze	13	81	0	12
3	Cloze	7	71	5.5	9
4	Editing	11	73	2.5	6
5	Editing	11	67	3.5	3.5
6	Editing	13	59	6	6.5
7	Editing	5	46	1.5	3
8	Editing	10	59	3	3

Note: LA= language analytical ability scores; LP= language proficiency scores;

Table 5.12 Mean and SD of language analytical ability and language proficiency scores

	LA	LP
Mean	10.00	62.50
SD	2.78	10.65

Note: LA= language analytical ability scores; LP= language proficiency scores;

Table 5.13 Mean and SD of two noticing scores

	Noticing 1	Noticing 2
Mean	3.13	5.94
SD	1.96	3.20

Table 5.14 All tests scores of the successful learners

Participant	Group	MC-pre	MC-post	MC-delayed	WO-pre	WO-post	WO-delayed
1	Non-output	0	4	5	5	17	24
2	Cloze	7	6	6	16	23	24
3	Cloze	5	4	9	4	8	12
4	Editing	7	8	6	0	3	15
5	Editing	6	4	5	11	21	30
6	Editing	8	8	8	15	21	24
7	Editing	6	5	1	0	3	6
8	Editing	6	7	4	5	21	22

Note: MC-pre= multiple-choice pretest; MC-post= multiple-choice posttest; MC-delayed= multiple-choice delayed-posttest; WO-pre= written output pretest; WO-post= written output posttest; WO-delayed= written output delayed-posttest;

Table 5.15 Mean and SD of multiple-choice test scores of successful learners

	MC-pre	MC-post	MC-delayed
Mean	5.63	5.75	5.50
SD	2.45	1.75	2.45

Note: MC-post= multiple-choice posttest; MC-delayed= multiple-choice delayed-posttest;

Table 5.16 Mean and SD of written output test scores of successful learners

	WO-pre	WO-post	WO-delayed
Mean	7.00	14.63	19.63
SD	6.28	8.55	7.89

Note: WO-pre= written output pretest; WO-post= written output posttest; WO-delayed= written

output delayed-posttest;

It can be seen from the above tables that successful learners from different groups seem to share a common character with each other. Their language proficiency and language analytical ability scores tend to be high. This might not be a surprising finding, but it is indicating the importance of individual differences in task performance and ultimate achievement in second language learning.

Further observation at the individual group level revealed that the only participants from the non-output group who made continuous improvement in throughout the two posttest stages appears to be highly grammatically sensitive and with high language proficiency. What is more interesting is that, while participants from this group tended to decline in their second noticing scores, her noticing score increased.

The two participants from the cloze group appears to have the highest language proficiency scores of all these successful learners, which seems to accord with what the overall data has found, that high proficient learners tend to benefit more from cloze task. Their language analytical ability scores are uneven, while one of them scored 13 out of total of 14 points; the other one scored only seven, which may indicate the less influential role of language analytical ability on cloze task.

On the other hand, the five successful learners from the editing group showed a tendency of high language analytical ability with only one exception. Their language proficiency though, appears to be not as high as the learners from the other two groups. Again, it seems to suggest the same thing as the overall data has revealed, that the editing task tends to have a facilitative effect on acquisition when it is carried out by the learners with high language analytical ability. Put in another way, learners with high language analytical ability tend to benefit more from editing task.

On the whole, the results of this small scale study are confirmative and consistent with those of overall data analyses. It is found that noticing does not happen to be the sufficient condition for learning to take place, but it is necessary. In order for noticed form to be converted into intake and thus acquired effectively, the language proficiency level and language analytical ability seems to have a role to play over the link. Set in the Gass's integrated model, in which she suggested that in order for comprehended input to be converted into intake, level of analyses made to the

comprehended input is crucial. Then, the findings from this study suggest that one way to facilitate this process is that the learners should be with higher language proficiency level and language analytical ability.

When it is concerned with the effect of task type on the link between noticing and acquisition, the findings of the qualitative analyses suggest that, the facilitative effect of the cloze task seems to be only of short term, and in a long run, editing task seems to be more effective in better promoting the acquisition.

Moreover, the findings also suggest that combination of a certain individual difference and a certain task type may better facilitate the acquisition. It seems the cloze task is in favor of the high proficient learners while editing task favors learners with high grammatical sensitivity. The same has been found form overall data analyses as well.

5.5 Summary of the chapter

This chapter presented the findings of the present study along with the discussion on those findings. The summary follows three main themes: the effect of output tasks on noticing of the target forms; the effect of output tasks on acquisition of the target forms; and lastly, the possible relationship between noticing and acquisition. Factors such as task type, individual differences and the possible role they play on noticing and acquisition, and the link between the two are also discussed.

5.5.1 Output and noticing

The first research question is concerned with whether output tasks could trigger the noticing of the target forms more than a non-output task, which does not require learners to produce the target language. The results of the present study showed that participants who carried out two output tasks in present study, the cloze reconstruction task and the editing reconstruction task respectively, noticed more target forms than those from the non-output group.

The finding is able to support the output hypothesis proposed by swain (1985; 1995; 2005). She claimed that by actually producing the target language, learners may have chance to notice the

gap between their IL and TL, and according to noticing hypothesis (Schmidt 1991, 2001) learners will only learn what they have noticed, but not much about what they have not.

The present study also set out to explore the possible role of task type and individual differences in influencing this output triggered noticing. As it turned out, both output tasks were equally effective in triggering noticing, although in a different ways due to the different nature of the two tasks. The missing information in the cloze task was obvious, thus, it could help learners to focus their attention to that specific parts. In order to carry out the editing task successfully, the learners needed to find out the erroneous part on their own, thus, those who noticed more errors tended to attend to the correct counterpart of those errors. Despite of their difference in nature, the two groups demonstrated same amount of noticing.

Another factor the present study aims to investigate is that, if learner inner qualities such as language proficiency or language analytical ability might play a role in influencing their ability to notice the target forms. The correlation analyses results showed that, language proficiency may have a role in affecting learner ability to notice, but not language analytical ability. Moreover, at individual group level, neither of them showed correlations with the noticing scores.

In sum, the findings suggest that output has a facilitative effect on triggering noticing of the target forms. However, output tasks may not differ that much when it concern with the amount of noticing. Furthermore, whether one is able to notice or not seems to have little to do with language analytical ability, but one may have a tendency to be influenced by his or her language proficiency level.

5.5.2 Output and acquisition

The second major research question deals with whether output tasks better facilitate the learning of the target forms compared to non-output task.

The results showed that between groups comparisons did not yield any evidence in favor of two output tasks. In other words, two output groups were expected to demonstrate a better performance on their test gains and consequently to outperform the non-output group. Statistical analyses revealed none such results, which meant output tasks failed to demonstrate a better facilitative effect on acquisition compared to the non-output task.

However, within group comparisons did provide mixed results. While the cloze group successfully improved on its production posttest scores, the editing group only showed a near significance tendency. Non-output, on the other hand, failed to make any significant improvement at all. This seems to indicate that although output tasks may not necessarily outperform non-output tasks, they did have a facilitative effect on the acquisition of the target forms.

This result is considered to be a partial support for the output hypothesis and noticing hypothesis. Learners form the cloze reconstruction group noticed the target forms more and therefore have improved in their production posttest. The editing group also demonstrated a similar tendency.

Another theoretical implication this finding suggest is that output task may only be facilitative in promoting the production skills of the learners, thus it may also support what the skill acquisition theory claims that, production practice only results in improvement in production skills, not in comprehension skills.

When the long term facilitative effect of output tasks was explored, it turned out that both output groups failed to extend their scores to the delayed-posttest stage. Thus, the long term effect of output tasks on learning was not confirmed according to the results of the present study.

The differential effect of Task type on learning of the target forms was not confirmed in the present study. The cloze task, despite showing facilitative effect on learning was not able to outperform the editing task at any test stage.

The results of correlation analyses and further grouping analyses about the effect of the individual differences however, provided a rather mixed findings. These findings seemed to suggest that a combination of a certain type of task and a certain individual difference may achieve a better facilitative effect on the acquisition. According to the findings of the present study, it seems that in order for learners to benefit from the cloze task, they should have a relatively high language proficiency level. Similarly, if the editing task is to be employed in the classroom, maybe only those with higher grammatical sensitivity may come up with a better performance.

5.5.3 Relationship between noticing and acquisition

The previous two sections showed that learners who noticed more may not necessarily perform better than those who noticed less. In other words, noticing may not always result in better acquisition. If it is so, is the opposite also possible? Put simply, if more noticing does not result in better acquisition, then better acquisition may also do not involve more noticing. Thus, a further analyses and qualitative observation are made to investigate the characters of the successful learners.

The results of above mentioned analyses showed that acquisition may not necessarily involve noticing, since the noticing scores of non-output group and editing group was relatively low. However, qualitative observation suggests that learners from cloze group which scored the highest noticing scores were likely to make improvement on their posttest scores.

Moreover, the further analyses of noticing scores of the editing group suggests that the low noticing scores did not necessarily mean that they did not notice. The number of attempts they made every time they carry out the task indicated that in fact they noticed more than they underlined. The noticing scores were low because they did not underline the whole passive structure, and that was due to the nature of the editing task. It appeared to be that half of the whole structure, either the verb *be* or past participle was already provided in the editing task sheet, and this made them to only underline half of the structure and resulted in low noticing scores. Since only 0.5 point is given to the underling of half structure.

Therefore, in fact, the majority of the learners who acquired the passive form successfully have actually noticed more.

Besides, the cloze task which triggered more noticing, it happened to be the one that also likely to promote acquisition of the form better. Though a statistical comparison did not support this claim, but almost half the number of participants from the cloze group managed to demonstrate improved performance while only five out of 23 from non-output group managed the same. The editing group demonstrated slightly better percentage than non-output group.

In addition, the mean scores of the language proficiency scores and language analytical ability test scores showed that these successful learners happened to share a similar character, that was, they were all more proficient and with higher grammatical sensitivity.

An interesting finding from this small scale study came from the qualitative observation made to see the long term effect of three tasks on learning. It was observed that in a longer term, the editing task seemed to have a more beneficial effect on learning, since five out of nine successful learners at the stage of posttest, also remained successful at the stage of delayed-posttest, while only one and two form the non-output group and the cloze group achieved the same. What is more, out of this five successful learners in editing group, four of them demonstrated relatively high language analytical ability scores, though not necessarily high language proficiency scores. This again, was consistent with findings from the previous section that maybe a certain task character in combination with a certain individual difference may lead to a better acquisition.

On the whole, the present study demonstrated that noticing, while not being the sufficient condition for learning to take place, certainly has its role in making it happen. Furthermore, this link between noticing and acquisition tends to be influenced by language proficiency level and the language analytical ability of the learners, and even sometimes, by the task character alone or combined with individual differences as mentioned above.

Chapter 6 Conclusion

6.1 Overview of the chapter

This chapter consists of following four parts. In the first part, the major findings of the present study are summarized. Then, the next section points out the limitations of the present study. In the following section, I propose suggestion for future studies. In final section, the pedagogical implications of the study are presented

6.2 Summary of the major findings of the study

The purpose of this study was to investigate the effects of output tasks on noticing and second language acquisition. The results of the study yielded following three major findings.

Firstly, the results supported that output tasks had a facilitative effect on triggering noticing of the target linguistic forms. The input-output-input-output treatment allowed the comparison of the changes took place between first underlining activity and second underling activity, which clearly showed the favorable effect of two output tasks on subsequent input compared to that of a non-output task.

Secondly, the effect of output tasks on acquisition of the target form was mixed. While the cloze reconstruction task enabled learners to improve their posttest performance on production test, the editing reconstruction task only showed a tendency of improvement. Moreover, the two output groups failed to outperform the non-output group at either posttest or delayed-posttest. Another finding concerning this research question was that, while output tasks showed a facilitative effect on learners' production skills, they did not show such effect on their comprehension skills.

Lastly, this study demonstrated that while noticing might not necessarily lead to acquisition, acquisition without noticing might not be possible. Besides, this possible link between noticing and acquisition might be influenced by factors such as task type, learners' proficiency level and aptitude to some extent.

6.3 Limitations of the study and future directions

Although the present study sheds light on output triggered noticing and possible relationship between noticing and acquisition, it is limited in number of ways. The first limitation is the generalizability of the findings due to the short nature of the treatment. Although the noticing function of the Output Hypothesis has been confirmed by this study, the possible facilitative effect of output was only partially confirmed. This might have been due to the short nature of the treatment, since each time learners were allowed only three minutes to read the input passage and five minutes to carry out the task. Since human attention is limited in its capacity when processing information, attention to one aspect of the task is achieved at the cost of neglecting other aspects (Anderson, 1995). Then it is possible that the completion of the task itself might have absorbed all their attention and left them little time for processing the noticed forms sufficiently enough to achieve a better acquisition. This may have been responsible for why two output groups while showing improvement in their production test scores still failed to outperform non-output group.

The second limitation is the operationalization of the language proficiency. Although some learners took part in present study were referred to as high proficiency group they were considered so only when they were compared to the low proficient group in this study. That means, learners in higher proficiency group were not necessarily advanced level learners. The mean scores of high groups were 70.02. This may also limit the generalizability of the findings of present study. Besides, one may question the validity of the language proficiency test used in this study. The language proficiency scores of the participants in present study have been based on their end-of-semester English test performance which was mainly on grammar knowledge. Furthermore, their scores showed remarkably high correlation with their college entrance examination test scores. College entrance examination is the official standardized test designed and administered by Ministry of Education of the People's Republic of China once a year to all the senior high school graduates. They will be selected by the colleges based on their performance on this test. Thus, the end-of semester test scores are considered as reliable in this study. However, it might be preferable for future studies that are interested in involving language proficiency as a variable to divide the groups based on some well known proficiency test scores such as TOEFL or

IELTS.

The third limitation is the operationalization of the task difference. One of the sub research questions in present study aimed at exploring the differential effects of two output tasks (i.e., cloze reconstruction task and editing reconstruction task) on noticing and learning of the target forms. Different operationalization of tasks may have different results. The two output tasks in present study were carried out in same format and mode (i.e., reconstruction task format and written form only). Future studies that are interested in task type effect should diversify the characters of the tasks so as to come up with a better understanding of the task differences. For example, as some previous studies have done, differential effects of individual tasks and collaborative tasks are worth further investigation (Nassaji & Tian, 2010; Storch, 2007); and differential effects of modality of tasks (i.e., written vs. oral) can also be explored (Izumi & Izumi 2004); or a same task can be operationalized into different level of complexity and difficulty following the triadic componential framework for L2 task design in Cognition Hypothesis proposed by Robinson (2001a). According to the Cognition Hypothesis, the increase in task complexity may (1) better facilitate greater accuracy and complexity in target language production; (2) heighten attention to, noticing of target forms made salient in input (see Robinson, 2005 for a review).

Furthermore, the qualitative observation in present study indicated the possibility of combining different task types with different individual differences might maximize the facilitative effect of the tasks on learning. For example, in present study the cloze task tends to be more effective in promoting learning when it is combined with learners with high language proficiency and the editing task seems to be more effective when it is carried out by learners with high language analytical ability. The similar suggestion has also been mentioned in the Cognition Hypothesis that the more complex and difficult the task becomes, the more likely individual differences in cognitive abilities (e.g., aptitude) or affective factors (e.g., motivation) would affect the outcome of the learning (Robinson, 2005).

The forth limitation is the inadequate measurement of the noticing. In this study underlining was implemented as the measurement of the noticing. Although underlining has its merit in providing online noticing, it is insufficient to provide closer look into the nature of link between noticing and acquisition. A combination of underlining with some follow up interview or think

aloud methods (see Robinson et al, 2012 for a review) might have provided a better understanding of nature of how the noticed forms processed by the learners. Maybe some learners just noticed and produced the forms to complete the task and upon completing the task just forgot the forms away. On the other hand, some of them may have undergone a process of comparing the noticed forms and their own IL and deepened their level of processing. Nevertheless, it should be noted that the think aloud techniques may have advantage in tapping into the more in depth nature of noticing, it also has its own limits. First, they may provide learners with extra opportunity to process the forms, thus confuse the effect of treatment and the technique itself. Second, learners may not be able to report everything they have noticed, especially when there is a time interval between treatment and the report. Thus, future studies need to consider the possibility of choosing two different but complementary methods to compensate the demerits of each. For example, Uggen (2012) combined underlining with stimulated-recall data collection and demonstrated that combination of quantitative and qualitative methods could provide "a much richer picture of learner internal process" (p. 531).

The fifth limitation is that this study did not include the comprehension check of the input passage by the two output groups. Lee (2007) found that textual enhancement (focus on form) promoted learning of target forms but had a negative effect on meaning comprehension. Thus, future studies that attempt to address the effect of output-oriented focus on form tasks may also need to explore the possible effect of such tasks on both learning of forms and meaning comprehension.

The sixth limitation is that, considering the overall low language proficiency of participants in present study, the choice of target structure (i.e., simple past tense in English passive voice) might have been a bit too challenging for them. It has been repeatedly pointed out by previous studies that despite the fact that structure of passive form is rather simple (verb *be* + past participle of verbs), L2 learners still cannot come up with a full mastery of the structure (Hinkel, 2002; Lee, 2007). A choice of different target structure might have produced different results.

6.4 Pedagogical implication

Despite the above mentioned limitations, the present study has provided some implications for L2

pedagogy. Firstly, two output tasks in present study have demonstrated their facilitative effect on promoting noticing of the target forms. Besides, Izumi (2002) has argued based on his findings that output tasks could trigger deeper level of processing of target forms and therefore promote learning of those forms. This makes output tasks as favorable candidates for L2 classroom focus on form activities.

Secondly, the study did not show the facilitative effect of output tasks on acquisition compared to non-output task. As has been discussed in previous section, this might have been due to the inadequate time spent on the input passage and task sheets, thus teachers should provide their students with sufficient time on both reading and output. Another option teachers can make is that they can ask students to work in pairs or groups. Nassaji & Tian (2010) reported the advantage of pair work on eliciting more focus on form talk between learners. This negotiation of form combined with effect of output may maximize the learning outcome of the students.

Thirdly, the present study has found that task features may require different individual differences to be involved to produce a more desirable outcome on learning (e.g., cloze task and language proficiency; language analytical ability and the editing task). This might be a message to the teachers that when choosing a pedagogical task they might as well consider the corresponding scaffolding ways to make sure the task can be utilized to its fullest possible effect on learning. For example, when language analytical ability of their students are not so high, they can make the certain linguistic rule more explicit or even provide a brief explanation of the rule underlying the structure.

Acknowledgements

First and foremost, I would like to express my deepest gratitude toward advising professors who have been most informative and helpful throughout my four-year doctoral study at Tohoku University: Prof. Kei Yoshimoto and Prof. Natsue Sugaya. Without their kind guidance, I surely would not have been able to complete my research. I am also grateful to Prof. Yoshio Kitahara for his valuable advice on the statistical analyses part of my dissertation, and Prof. Wataru Nakamura for his comments on my research.

Special thanks are due to Prof. Nobuya Itagaki, Prof. Wataru Suzuki and Adrian Leis from Miyagi University of Education for their continued help both academically and privately throughout my master's and doctoral study. Without their guidance and help, I would not have been able to continue my study.

There are also various individuals that contributed to the preparation stage of the experiment for this paper in a number of ways whom I would like to thank here. Thank you to Dr. Yoko Mukoyama for your insightful comments on the research design of the experiment. Thank you to Takashi Oba for recommending and providing the input passage for the experiment. Thank you to Prof. Nina Spada from Toronto University for kindly agreeing to the full usage of the written output test materials from her study. Thank you to Adrian Leis for the proof reading of all materials as well as your kind encouragement. Thank you to Konomi Ariji and Moe Hashiba for help and support during the preparation stage of the testing and treatment materials. Thank you to Xiao Feng from Inner Mongolia University of Finance and Economics for allowing me to conduct my research in her class and her lovely students for sacrificing their class time to participate in my research.

I am also grateful to all my friends both from Inner Mongolia and Japan, my fellow researchers and students in the graduate school of International Culture studies for their helpful comments and informative talks throughout my time there.

Last but not least, I would like to thank my family for their love and support, which helped me through the hard times to strive towards my goal.

References

向山陽子 (2009) 「第二言語習得において学習者の適性が学習成果に与える影響―言語分析能力・音韻的短期記憶・ワーキングメモリに焦点を当てて―」『日本語科学』25, 67-90

向山陽子 (2013) 『第二言語習得における言語適性の役割』ココ出版

Borjigin, S., Suzuki,W., & Itagaki, N. (2013). The Effect of An Output Task on L2 Acquisition: Does task difficulty make a difference? *Research Bulletin of Tohoku English Language Education Society, 33,* 59-72

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*, 195-221.

Dörnyei, Z. & Ushioda, E. (2010). *Teaching and Researching Motivation*, (2nd ed.). Harlow: Longman.

Ellis, R. (2003). *Task-based language teaching and learning.* Oxford: Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.

Garica Mayo, M. P. (2002b). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics*, *12*, 156-175.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9*, 198-217.

Ghari, A., & Moinzadeh, A. (2011). The effects of output task types on noticing and learning of English past models: a case of intermediate Persian adult learners of English. *Journal of Language Teaching and Research,2,* 1180-1191

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four- stage writing task. *Language Teaching Research, 11*, 459-479.

Hinkel, E. (2002). Why English passive is difficult to teach (and learn). In Eli. Hinkel, & Sandra. Fotos (Eds), *New perspective on grammar teaching in second language classrooms* (pp. 233-259). London: Lawrence Erlbaum Associates.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24,* 541-577.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24,* 168-196.

Izumi, S, & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34,* 239-278.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21,* 421-452.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman. London: Laurence Erlbaum Associates.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge

Lee, S-K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning, 57,* 87-118.

Lee,S-K., & Huang, H-T. (2008). Visual input enhancement and grammar learning: a meta-analytic review. *Studies in Second Language Acquisition, 30,* 307-331.

Mackey,A., & Sachs,R,. (2012). Older Learners in SLA Research: A First Look at Working Memory, Feedback, and L2 Development. *Language Learning, 62,* 704–740.

Mackey, A., Philp. J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning*. (pp. 181-210). Amsterdam: Benjamins.

Muranoi, H. (2007). Output practice in the L2 classroom. In Robert, DeKeyser (Ed.), *Practice in second language learning: Perspective from applied linguistics and cognitive psychology* (p. 51-84). Cambridge University Press.

Nassaji, H., & Tian, J. (2010). Collaborative and individual task and their effects on learning English phrasal verbs. *Language Teaching Research, 14,* 397-419.

Ortega, L. (2009). Understanding second language acquisition. London: Hodder Education

Qi, D., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10,* 277-303.

Ranta, L. (1998). *The significance of grammatical sensitivity for L2 learning in communicative ESL classrooms* (Unpublished doctoral dissertation). Concordia University, Montreal.

Robinson, P. (1995). Attention, memory and the “noticing” hypothesis. *Language Learning, 45,* 283-331.

Robinson, P. (1997a). Individual differences and the fundamental similarity of implicit and explicit second language learning. *Language Learning, 47,* 45-99.

Robinson, P. (1997b). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition, 19,* 223-247.

Robinson,P., Mackey,A., Gass, S., & Richard Schmidt. (2012). Attention and awareness in second language acquisition. In Susan Gass and Alison Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition*, (pp. 247-267). New York: Routledge.

Sawyer, M,. & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In Peter Robinson (ed.), *Cognition and second language instruction.* UK: Cambridge Univerisity press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11,* 129-158.

Schmidt, R. (2001). Attention. In Peter, Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, N., Dörnyei ., Z., Adolphs, S., & Durow, V. (2003). Knowledge and acquisition of formulaic sequences: A longtitudinal study. In N. Schmitt (Ed.), *The acquisition , processing, and use of formulaic sequences* (pp. 55-86). Amsterdam: John Benjamins.

Sheen, Y. (2007a). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41,* 255-283.

Sheen, Y. (2007b). The effect of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 301-322). Oxford: Oxford University Press.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36,* 1-14

Song, M-J., & Suh, B-R. (2008). The effect of output task types on noticing and learning of the English past counterfactual conditional. *System, 36*, 295-312.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 30*, 73-87.

Spada, N. (2011). Beyond form-focused instruction: Reflection on past, present and future research. *Language Teaching, 44, 2*, 225-236.

Spada, N., Jessop, L., Tomita, Y., Suzuki, W. & Valeo, A. (2014). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research, 18*, 453–473.

Steven, J.M., & Bliss, B. (2003). *Side by Side Book 4* (3rd ed.). New York: Pearson Longman.

Storch, N. (1997). The editing talk of adult ESL learners. *Language Awareness, 6*, 221–32.

Storch, N. (2007). Investigating the merits of pair work on a text editing task in ESL classes. *Language Teaching Research, 2*, 143–59.

Suzuki, W., Itagaki, N., Takagi, T., & Watanabe, T. (2009). The effect of output processing on subsequent input processing: a free recall study. Teachers College, Columbia University *Working Papers in TESOL &Applied Linguistics, 9*, 1-17.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Susan. Gass, & Carolyn. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In Guy, Cook. & Barbara, Seidelhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In Eli. Hinkel (Ed.), *The hand book of second language teaching and learning* (pp. 471-483). Mahwah, NJ: Lawrence Erlbaum.

Swain, M., & Lapkin, S. (1995). Problems in output and cognitive processes they generate: A step towards second language learning. *Applied linguistics, 16*, 371-391.

Uggen, S.M. (2012). Reinvestigating the noticing function of output. *Language Learning, 62*, 506–540.

Wong, W. (2003). Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning, 13,* 17-45.

Appendices

Appendix 1 Testing materials

1.1 Language aptitude test

学号: _____

Language Analysis

The list in the box below contains words/phrases from an imaginary language along with their English translation. Following this, there will be 14 short English sentences, each with four possible translations into the imaginary language. Based on the examples given in the box, we would like to ask you to try and work out which of the four options is the correct translation of each sentence. Thank you very much.

kau: dog

meu: cat

kau meud bo: The dog is chasing the cat.

kau meud bi: The dog was chasing the cat.

so: watch

ciu: mouse

pa: we, us

xa: you

pasau meud bo: Our dog is chasing the cat.

pa meud bo: We are chasing the cat.

paxbo: We are chasing you.

pa meud bor: We aren't chasing the cat.

1. **The dog is watching the cat.**
   A. kau meud so **B.** kau meud si
   C. meu kaud so **D.** meu kaud si

2. **The cat was watching the mouse.**
   A. meud ciu so **B.** meu ciud so
   C. meud ciu si **D.** meu ciud si

3. **You are watching us.**
   A. paxbo **B.** paxso
   C. xapbo **D.** xapso

4. **You were chasing the dog.**
   A. xa kaud bo **B.** pa kaud bo
   C. pa kaud bi **D.** xa kaud bi

5. **We were watching you.**
   A. xapsi **B.** paxso
   C. paxsi **D.** paxbi

6. **You are not watching the cat.**
   A. xa meud bor **B.** xa meud sor
   C. xa meud sir **D.** xa meu sor

7. **You are not chasing us.**
   A. paxbor **B.** xapbo
   C. xapabor **D.** xapbor

8. **We were not watching the dog.**
   A. pa kaud sir **B.** pa kau sir
   C. pa kaud sor **D.** pa kaud bir

9. **We were not chasing you.**
   A. xapbir **B.** paxbir
   C. paxbor **D.** xapbor

10. **Your cat is chasing the mouse.**
    A. xacu meud bo **B.** xaseu ciud bo
    C. meuxa ciud bo **D.** ciuxa meud bo

11. **You are not watching our dog.**
    A. xa paseud bor **B.** xa pasaud sor

C. xa pasaud so **D.** xa pasaud bor

12. Our mouse was not chasing the dog.

A. oasiu kaud bi **B.** xasiu kaud sir
C. xasiu kaud bi **D.** pasiu kaud bir

13. Your mouse is chasing us.

A. xa ciu pabo **B.** xasiu pbo
C. xaciu pa bo **D.** xasiu pabo

14. Our cat was not chasing your dog.

A. pseu xasaud bir **B.** pseu xsaud bir
C. paseu xasaud bir **D.** paseu xsaud bir

1.2 Multiple-choice tests

1.2.1 Pretest and delayed-posttest

学号_____

语法测试

从 A,B,C,D 中选择最恰当的选项，写在每道题前面的括号内。

( ) 1. If you had come a bit earlier, you _____ him.

A. may see B. may have seen C. may be seen D. may be seeing

( ) 2. The staff of our company _____ there by plane.

A. was going to B. is going C. are going D. were going to

( ) 3. A police officer_____in a car accident last night.

A. injured B. had injured C. was injuring D. was injured

( ) 4. _____ to the left, you'll find a tall hotel building.

A. To turn B. Turning C. Turns D. Turned

( ) 5. If my grandmother_____alive, she would have been eighty next year.

A. were B. was C. had been D. has been

( ) 6. What the native speaker said_____by the students because their English was so poor.

A. did not understand B. was not understanding

C. was not understood D. was not understand

( ) 7. He or I _____ going to the cinema.

A. is B. am C. are D. be

( ) 8. If we____a map, we would find the place easily

A. have B. had C. are having D. can have

( ) 9. It_____that the earth was the center of the universe.

A. was believing B. was believed C. believed D. believes

( ) 10. I haven't seen him since he _____.

A. is leaving B. leaves C. has left D. left

( ) 11. I_____in a difficult situation if Jane hadn't helped me.

A. would be B. was C. would have been D. had been

( ) 12. The house_____when they moved in last year.

A. repainted B. was repaint C. was repainting D. was repainted

( ) 13. I_____hard until I pass the TOEFL.

A. study B. have studied C. would study D. will study

( ) 14. By the end of this weekend, we_____the work.

A. already finished B. will already finish

C. will be already finished D. will have already finished

( ) 15. When we came back from the party, we found that our car_____.

A. was stealed B. was stoled C. was stealing D. was stolen

( ) 16. I received a letter from my sister the other day ____that she is coming to visit me next month.

A. says B. saying C. said D. have said

( ) 17. How I wish I _____ your advice!

A. should take　　B. take　　C. took　　D. had taken

(　) 18. Yesterday I____by Mr. Jones to attend the meeting although I was not so willing to go.

A. asked　　B. was asking　　C. was asked　　D. was ask

(　) 19. If only it_____we could go out for picnic.

A. stops to rain　B. had stopped to rain　C. would stop raining　D. stopped raining

(　) 20. A lot of people in Japan spend their leisure time____pachinko.

A. playing　　B. to play　　C. played　　D. play

(　) 21. I_____to the party, but I went anyway.

A. did not invite　　B. was not inviting　　C. was not invited　　D. did not invited

(　) 22. The train_____when we hurried to the station.

A. already left　　B. was already left　　C. has already left　　D. had already left

(　) 23. If I were a rich person, I_____travel around the world.

A. will　　B. shall　　C. can　　D. would

(　) 24. The new classroom building_____last month.

A. completed　　B. was completing　　C. had completed　　D. was completed

(　) 25. When you go out, please remember to close your bedroom window _____.

A. if it will rain　　B. unless it rains　C. in case it rains　　D. whether it will rain

(　) 26. We'll go hiking _____ it rains or shines tomorrow.

A. if　　B. whether　　C. since　　D. though

( ) 27. The famous actor and actress_____walking along the beach together last Sunday.

A. saw B. were seeing C. were seen D. have seen

( ) 28. If he_____more careful, he would not have made such a stupid mistake.

A. was B. were C. has been D. had been

( ) 29. Please keep the receipt for your new dress. Otherwise, you _____be able to exchange it.

A. should not B. could not C. might not D. must not

( ) 30. We went to visit the museum last week, but it_____. So we went to the cinema instead.

A. closes B. has closed C. was closed D. was closing

1.2.2 Posttest

学号_____

语法测试

从 A,B,C,D 中选择最恰当的选项，写在每道题前面的括号内。

( ) 1. The tallest building in this city_____in the earthquake last week.

A. destroyed B. had destroyed C. was destroying D. was destroyed

( ) 2. If Jane had come a bit earlier, she _____ him.

A. might see B. might have seen C. might be seen D. might be seeing

( ) 3. _____ to the left, you'll find a tall hotel building.

A. To turn B. Turning C. Turns D. Turned

( ) 4. What the coach said_____by the skater because it was very noisy.

A. did not hear B. was not hearing C. was not heard D. was not heared

( ) 5. It is better _____ than to cry.

A. laugh B. to laugh C. to be laughed D. laughing

( ) 6. If Aunt Mary_____alive, she would have been eighty next year.

A. were B. was C. had been D. has been

( ) 7. It _____that some people in this area saw a UFO in the night sky.

A. was reporting B. was reported C. reported D. reports

( ) 8. I haven't seen him since he _____.

A. is leaving B. leaves C. has left D. left

( ) 9. I _____ hard until I pass the TOEFL.

A. study B. have studied C. would study D. will study

( ) 10. Sally lost her bike last month, but two days ago it_____to her secretly.

A. has returned B. was returned C. returned D. was return

( ) 11. I_____in a hard situation if my teacher hadn't helped me.

A. would be B. was C. would have been D. had been

( ) 12. I received a letter from my sister the other day_____that she iscoming to visit me next month.

A. says B. saying C. said D. have said

( ) 13. When I went to the office, I found that the window glass_____.

A. was breaking B. broked C. was broked D. was broken

( ) 14. Ann did well in the final exam, since she_____hard recently.

A. works B. is working C. has worked D. has been working

( ) 15. The car race would have begun at eight in the morning, if it _____ started raining.

A. was not B. did not C. has not D. had not

( ) 16. The dancer_____by the audience to perform one more dance and she did.

A. requested B. was requesting C. was requested D. was request

( ) 17. We_____dinner when somebody knocked at the door.

A. had B. are having C. were having D. have had

( ) 18. How I wish I_____your suggestion!

A. should take　　B. take　　C. took　　D. had taken

( ) 19. I_____to the house warming party of my new neighbor, but I went anyway.

A. did not invite　　B. was not inviting　　C. was not invited　　D. had not invited

( ) 20. What_____be doing this time tomorrow?

A. do you　　B. are you　　C. is you　　D. will you

( ) 21. The train_____when we hurried to the station.

A. already left　　B. was already left　　C. has already left　　D. had already left

( ) 22. Two boys _____ to have been absent from the class in the morning.

A. knew　　B. were knowing　　C. were known　　D. known

( ) 23. If only it_____we could go out to play soccer.

A. stops to rain　　B. had stopped to rain　　C. would stop raining　　D. stopped raining

( ) 24. They _____living in that house for fifty years when they decided to move.

A. had been　　B. have been　　C. are　　D. was

( ) 25. Her new book_____into Chinese last year and since then it has been the most popular book on the list.

A. translated　　B. has translated　　C. was translating　　D. was translated

( ) 26. Drinking too much will do harm_____your health.

A. in　　B. with　　C. to　　D. on

( ) 27. If I were you, I_____say hello to him.

A. will　　B. shall　　C. can　　D. would

( ) 28. The document _____and everything was ready for the meeting.

A. completed B. was completing C. has completed D. was completed

( ) 29. He behaves as though he _____the professor well.

A. knew B. know C. is knowing D. known

( ) 30. If he_____more hard working, he would not have been this poor.

A. was B. were C. has been D. had been

Appendix 2 Treatment packages

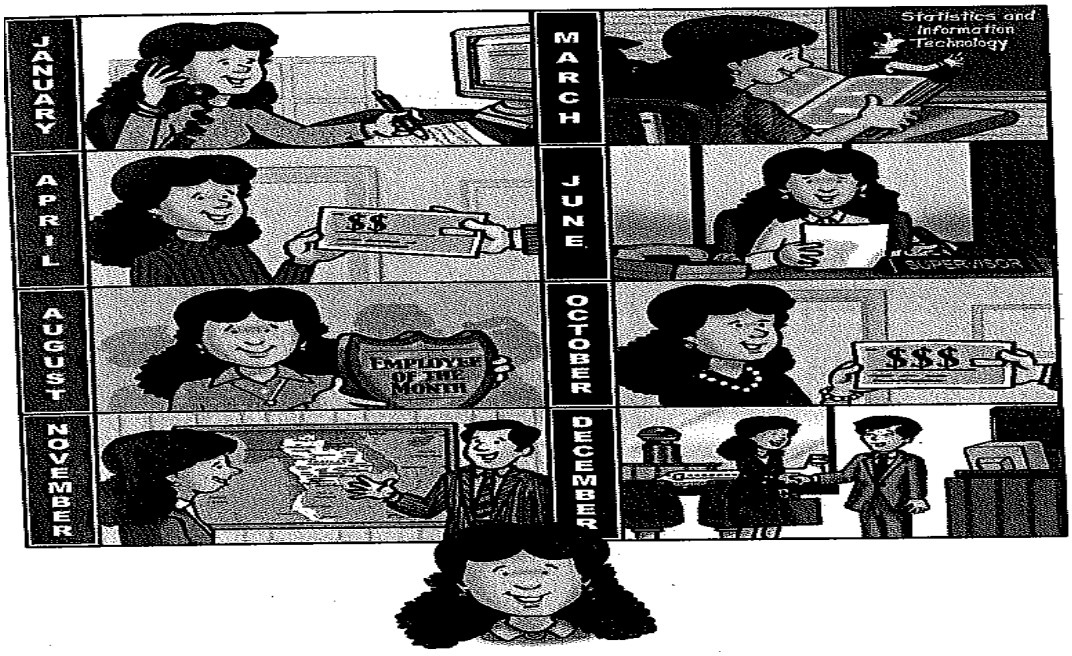
2.1 Non-output task package

封面

学号_____

没有老师的指示之前请不要翻开此页！

阅读下面的短文，并在你认为重要的词或短语下面画线。(3 分钟)



A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

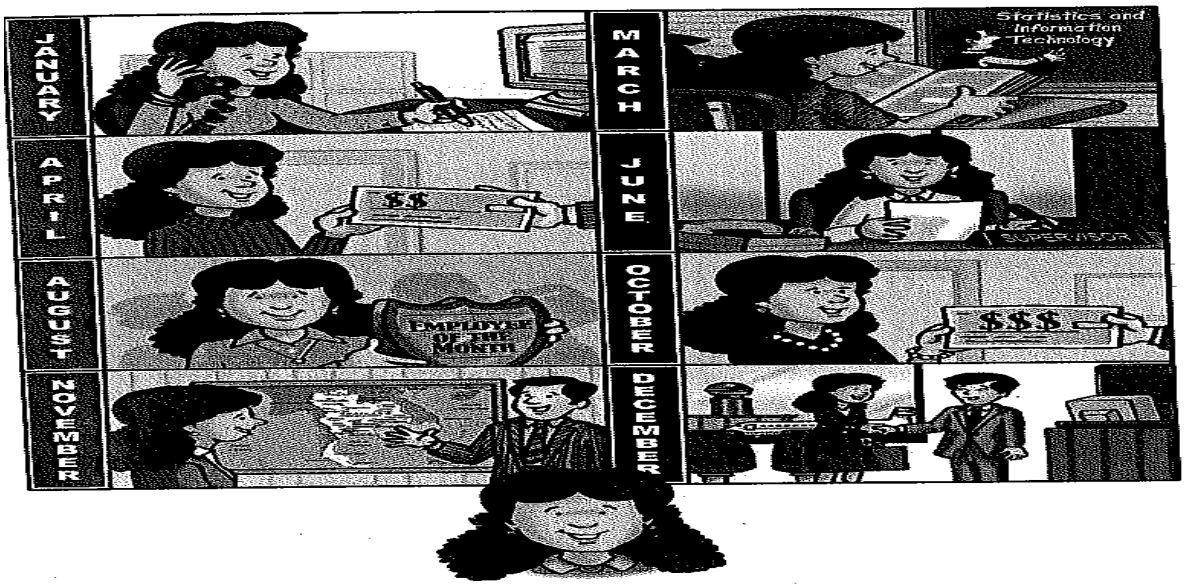
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

<u>没有老师的指示之前请不要翻开此页!</u>

请不要翻回前一页!

根据短文判断下列句子是否正确。请在每个句子前面的括号里填写√或×。(5 分钟)



( ) 1. Amy is a sales woman in a telephone company.

( ) 2. Amy went to school to study foreign languages, because the company asked her to do so.

( ) 3. Amy became the supervisor of her department in July.

( ) 4. The company has given raises to her three times within a year.

( ) 5. Amy wanted to apply for the new position but the company said no.

( ) 6. Amy quitted her job and moved to Bangkok, Thailand to get a new job.

( ) 7. The company awarded Amy the "Employee of the Month" twice.

( ) 8. Amy had a very exciting year, because she had many chances to go abroad to company's overseas offices.

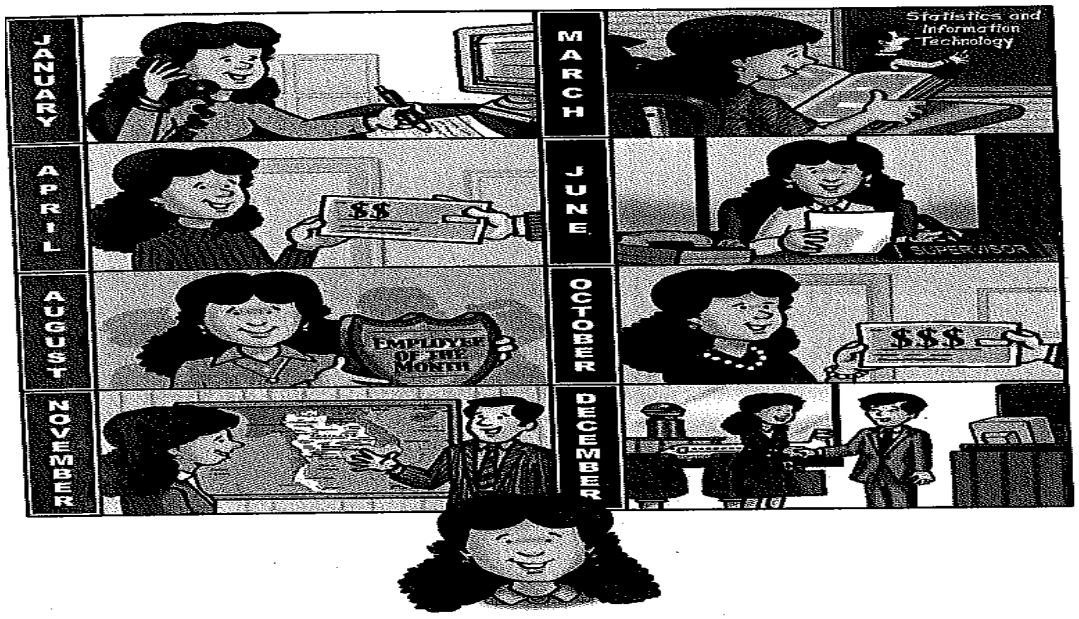
( ) 9. Amy went to Thailand in November.

( ) 10. Since her employment in January, Amy has had many exciting experiences at work.

没有老师的指示之前请不要翻开此页!

请不要翻回前一页!

阅读下面的短文，并在你认为对接下来的判断正误题重要的词或短语下面画线。(3 分钟)



A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

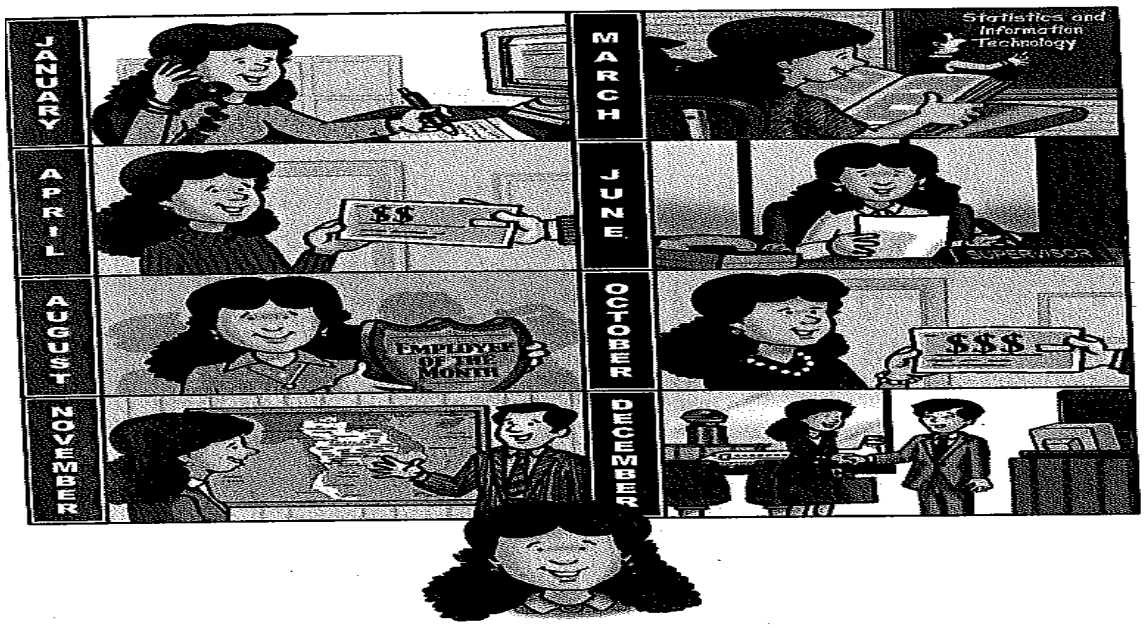
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

没有老师的指示之前请不要翻开此页！

请不要翻回前一页!

根据短文判断下列句子是否正确。请在每个句子前面的括号里填写√或×。(5 分钟)



A very exciting year

( ) 1. Amy is a sales woman in a telephone company.

( ) 2. Amy went to school to study foreign languages, because the company asked her to do so.

( ) 3. Amy became the supervisor of her department in July.

( ) 4. The company has given raises to her three times within a year.

( ) 5. Amy wanted to apply for the new position but the company said no.

( ) 6. Amy quitted her job and moved to Bangkok, Thailand to get a new job.

( ) 7. The company awarded Amy the "Employee of the Month" twice.

( ) 8. Amy had a very exciting year, because she had many chances to go abroad to company's overseas offices.

( ) 9. Amy went to Thailand in November.

( ) 10. Since her employment in January, Amy has had many exciting experiences at work.

谢谢合作! ☺

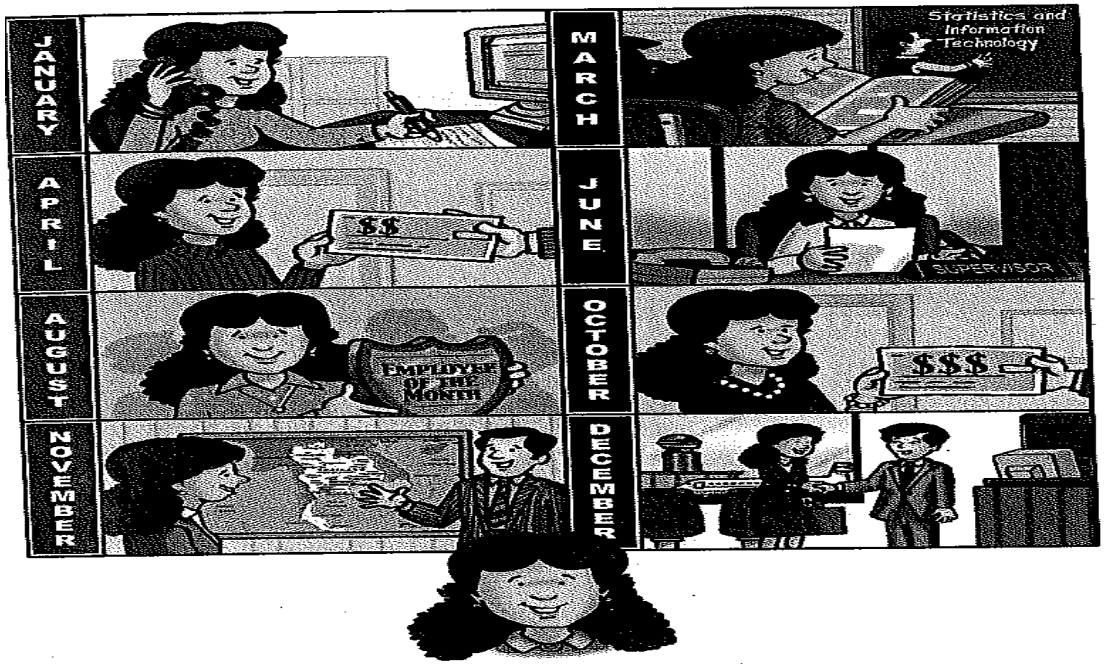
2.2 Cloze reconstruction task package

封面

学号_____

没有老师的指示之前请不要翻开此页！

阅读下面的短文，并在你认为重要的词或短语下面画线。(3 分钟)



A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

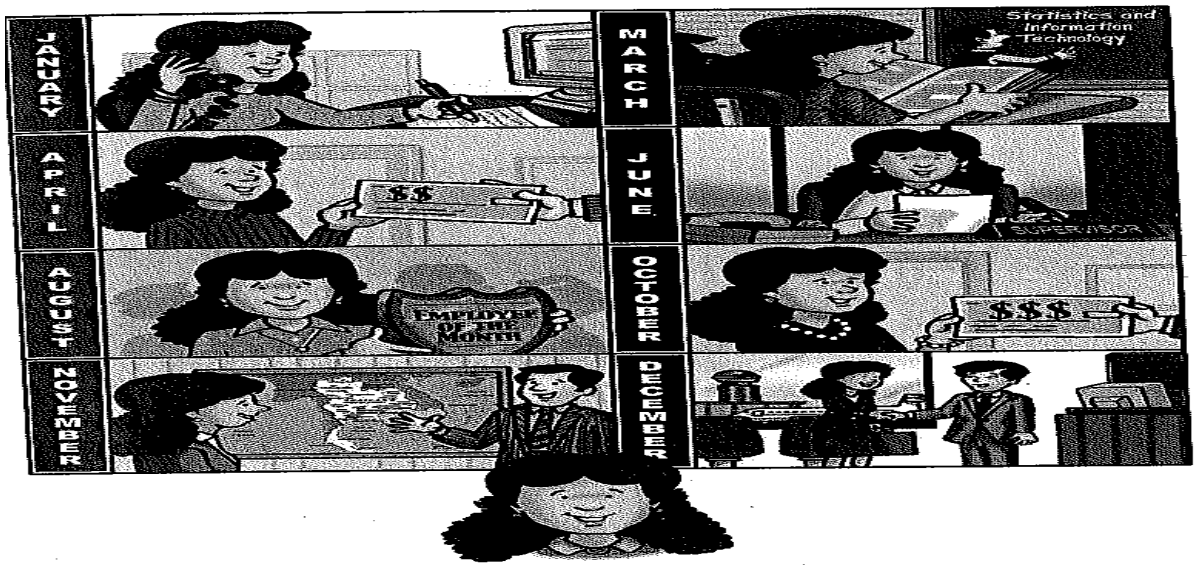
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

<u>没有老师的指示之前请不要翻开此页！</u>

请不要翻回前一页!

根据你刚度过的短文的内容，填空。(5 分钟)



A very exciting year

In January, Amy (　　) as (　　) secretary (　　) the Inter-Tel company.

In March, she (　　) to school (　　) the company to study statistics and information technology.

In April, she (　　) a raise.

Just two months (　　), she (　　) to the position of supervisor of her department.

In August, she (　　) as the “Employee of the Month”.

In October, she (　　) another raise.

In November, she (　　) to apply for a position in the company’s overseas office (　　) Bangkok, Thailand.

In December, she (　　) the new position (　　) Thailand.

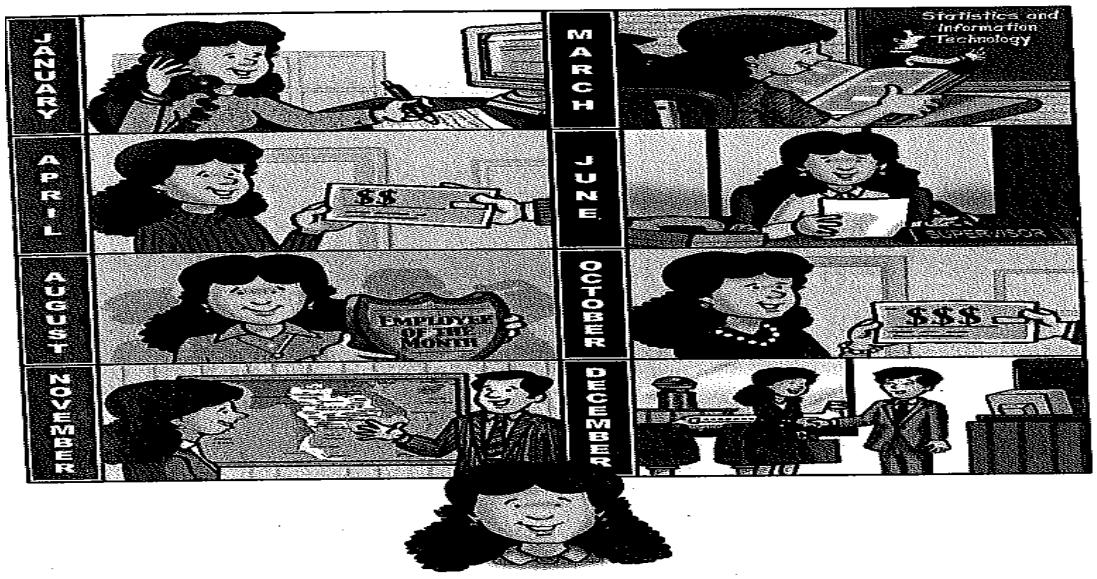
At (　　) month, she (　　) the “Employee of the Year”.

She can’t believe all the wonderful (　　) that (　　) to her since she (　　) just twelve months ago.

没有老师的指示之前请不要翻开此页！

请不要翻回前一页!

阅读下面的短文，并在你认为对接下来的填空题重要的词或短语下面画线。(3 分钟)



A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

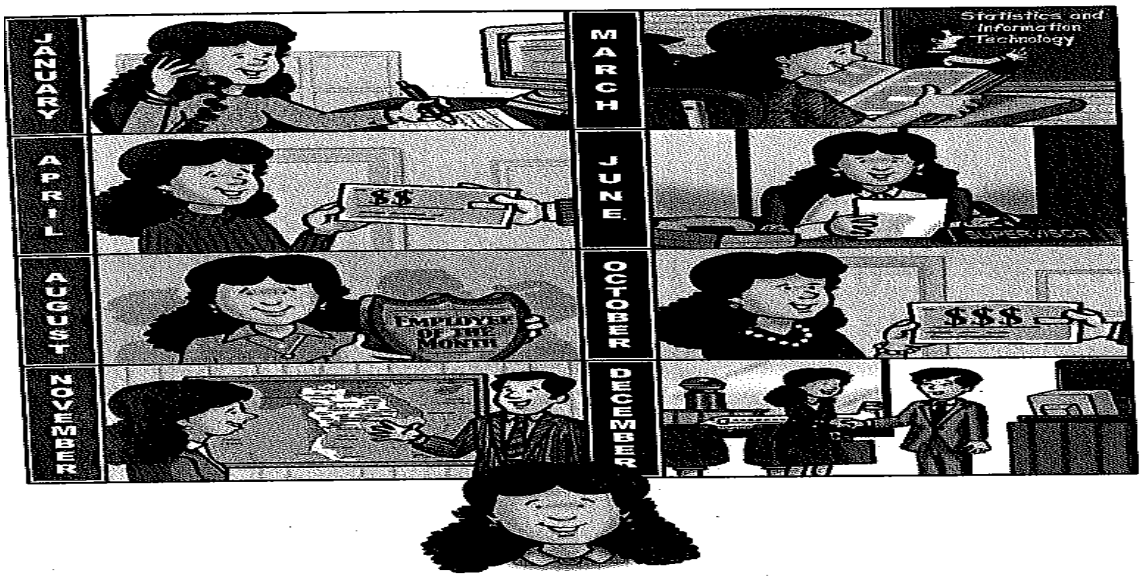
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

没有老师的指示之前请不要翻开此页！

请不要翻回前一页!

根据你刚度过的短文的内容，填空。(5 分钟)



A very exciting year

In January, Amy (　　) as (　　) secretary (　　) the Inter-Tel company.

In March, she (　　) to school (　　) the company to study statistics and information technology.

In April, she (　　) a raise.

Just two months (　　), she (　　) to the position of supervisor of her department.

In August, she (　　) as the "Employee of the Month".

In October, she (　　) another raise.

In November, she (　　) to apply for a position in the company's overseas office (　　) Bangkok, Thailand.

In December, she (　　) the new position (　　) Thailand.

At (　　) month, she (　　) the "Employee of the Year".

She can't believe all the wonderful (　　) that (　　) to her since she (　　) just twelve months ago.

谢谢合作! 

2.3 Editing reconstruction task package

封面

学号_____

没有老师的指示之前请不要翻开此页！

阅读下面的短文，并在你认为重要的词或短语下面画线。(3 分钟)

A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

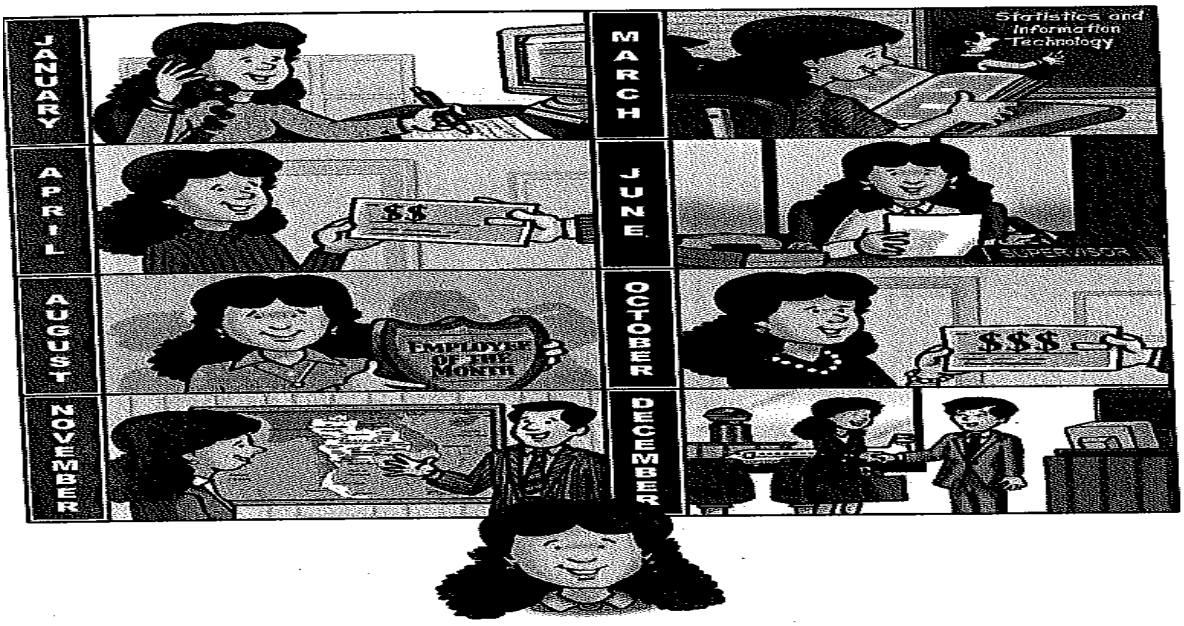
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

<u>没有老师的指示之前请不要翻开此页!</u>

请不要翻回前一页!

根据刚读过的短文，从以下句子中找出有错误的地方，并改正。每个句子包括一个或一个以上的错误。(5 分钟)



A very exciting year

In January, Amy employed as the secretary for the Inter-Tel company.

In March, she sended to school with the company to study statistics and information technology.

In April, she was giving a raise.

Just two months late, she promoted to the position of supervisor of her department.

In August, she chooses as the “Employee of the Month”.

In October, she gave another raise.

In November, she was inviting to apply for a position in the company’s overseas office on Bangkok, Thailand.

In December, she provided the new position of Thailand.

At end of month, she was award the “Employee of the Year”.

She can’t believe all the wonderful thing that has happened to her since she was hire just twelve months ago.

没有老师的指示之前请不要翻开此页！

请不要翻回前一页!

阅读下面的短文，并在你认为对接下来的改错题重要的词或短语下面画线。(3 分钟)

A very exciting year

In January, Amy was employed as a secretary by the Inter-Tel company.

In March, she was sent to school by the company to study statistics and information technology.

In April, she was given a raise.

Just two months later, she was promoted to the position of supervisor of her department.

In August, she was chosen as the “Employee of the Month”.

In October, she was given another raise.

In November, she was invited to apply for a position in the company’s overseas office in Bangkok, Thailand.

In December, she was provided the new position in Thailand.

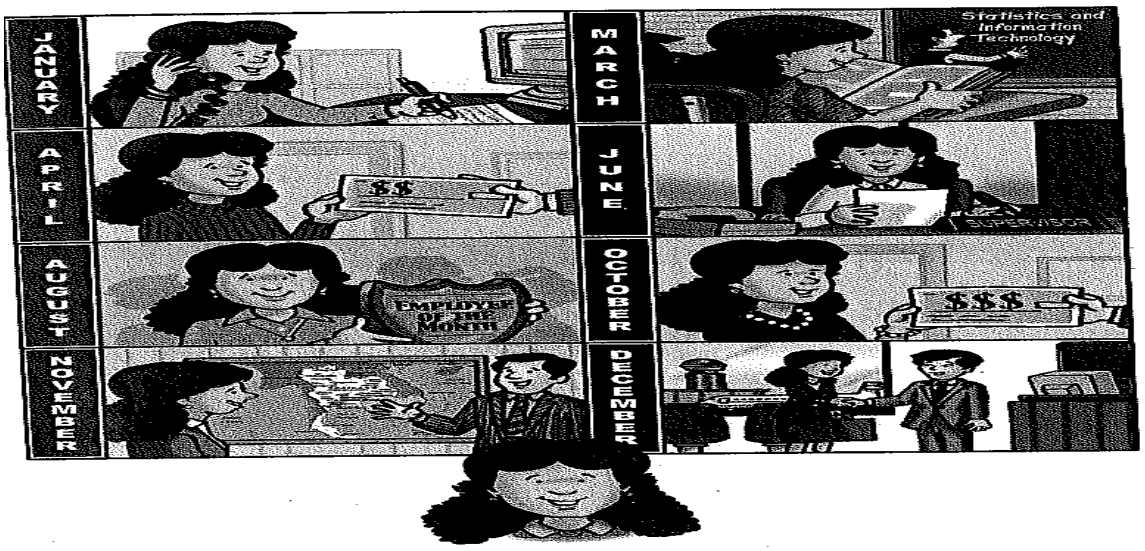
At the end of the month, she was awarded the “Employee of the Year”.

She can’t believe all the wonderful things that have happened to her since she was hired just twelve months ago.

没有老师的指示之前请不要翻开此页！

请不要翻回前一页!

根据刚读过的短文，从以下句子中找出有错误的地方，并改正。每个句子包括一个或一个以上的错误。(5 分钟)



A very exciting year

In January, Amy employed as the secretary for the Inter-Tel company.

In March, she sended to school with the company to study statistics and information technology.

In April, she was giving a raise.

Just two months late, she promoted to the position of supervisor of her department.

In August, she chooses as the “Employee of the Month”.

In October, she gave another raise.

In November, she was inviting to apply for a position in the company’s overseas office on Bangkok, Thailand.

In December, she provided the new position of Thailand.

At end of month, she was award the “Employee of the Year”.

She can’t believe all the wonderful thing that has happened to her since she was hire just twelve months ago.

谢谢合作! ☺

Appendix 3 Treatment procedure instructions manuscript (Chinese version)

实验要领

1. 口头传达以下内容。

a. 从现在开始要做的调查（包括考试和练习）所得的分数，跟大家学校里的学习成绩无关，不会对此产生任何影响，敬请放心解答。
b. 请大家填写自己的学号，但不用填写名字。
c. 大家此次成绩只为研究目的所使用，个人信息将受到严密的保护。
d. 从现在开始发带有表纸的册子，请听从老师的指示，不要擅自翻开任何一页。

2. 发放册子时注意表纸一定要朝上。

「从现在开始发册子，请填好自己的学号。还没有老师的知识的情况下请不要擅自翻页。」

实验实施中的指示

3. 参加实验者将阅读并背诵短文。

「请大家翻开表纸，翻到第一页，请仔细阅读指示并按照指示背诵短文。规定时间为 3 分钟。」

「还剩 30 秒。」

4. 拿 A 类册子的试验者做填空题，B 类的做改错题，C 类的做选择题。

「时间到，请翻开下一页，反倒第二页，仔细阅读并根据指示做。规定时间为 5 分钟。开始。」

「还剩 1 分钟。」

5. 参加实验者重读短文。

「时间到。翻到下一页。翻到第三页。请仔细阅读指示并按照指示背诵短文。规定时间为 3 分钟。」

「开始。」

「还剩 30 秒。」

6. 拿 A 类册子的试验者做填空题，B 类的做改错题，C 类的做选择题。

「时间到，请翻开下一页，反倒第四页，仔细阅读并根据指示做。规定时间为 5 分钟。」

「开始。」

「还剩 1 分钟。」

7. 回收册子。

「时间到，请停止作答。现在开始回收册子，请将表纸朝上放在桌子上。最后再次确认有没有填写学号。」

「好，回收册子。」

8. 进行语法考试。

「现在进行语法考试。开始发卷子，还没有老师的指示的情况下，不要擅自开始作答。」

「好，开始做题。规定时间为 20 分钟。」

「还剩 1 分钟。」

「好，时间到。开始收卷子。」

9. 看图写作文。先发作文的答题纸。

「现在开始看图写作考试。分到答题之后，请仔细阅读指示，并等待老师的指示。」

「好，现在开始放幻灯片。请看图片，并根据答题纸上的提示词，描述这幅图。每一幅图的展示时间为 （1） 分钟。」

「好，最后一幅图展示完毕。」
「请最后再次检查有没有填好学号。好，收答题纸」

10. 实验结束。

「调查结束。感谢你们的配合！谢谢！

Appendix 4 Background information questionnaire

根据您的实际情况回答下列问题。请在相应的“□”上打钩。

1. 您的性别:

□男 □ 女

2. 您的年级:

□大一 □大二 □大三 □大四

3. 您的专业是：(　　)

4. 您开始学习英语的时期是:

□小学以前 □从小学开始 □ 从中学开始 □ 从高中开始 □进入大学以后

□ 其他（请注明：　　）

5.请选择您参加过的国家级英语考试。并请在括号里填写您所获得的分数（请注明参加考试的年月份)。

□英语三级(_____年_月 分数为____)

□英语四级(_____年_月 分数为____)

□英语六级(_____年_月 分数为____)

其他(考试名称_____ _____年_月 分数为____)

□没有参加过

6.进入大学以后您所在的英语班上所使用的教学语言为:

□中文和英语 □蒙古语和英语 □ 蒙古语，中文和英语 □ 只有英语

7.您从哪里来？（　　）盟（　　）市 （　　）旗县（　　）苏木